Noname manuscript No.
(will be inserted by the editor)

A combined first and second order variational approach for image reconstruction

Konstantinos Papafitsoros · Carola-Bibiane Schönlieb

February 29, 2012

Abstract In this paper we formulate and study a variational problem in the space of functions of bounded Hessian $BH(\Omega)$. This forms a higher order extension of the well known ROF functional (total variation minimisation) which is widely used for image reconstruction. Our functional involves convex functions of the total variation and the total variation of the first derivatives. We prove existence and uniqueness of minimisers via the method of relaxation. We use the split Bregman method in order to solve numerically the corresponding discretised problem and we prove some convergence results. We apply our model to image denoising (Gaussian and impulse noise) and deblurring as well as image inpainting. The numerical results show that our model avoids the creation of undesirable artifacts and blocky-like structures in the reconstructed images which is a disadvantage of the ROF model.

Keywords Functions of bounded Hessian · Split Bregman · Denoising · Deblurring · Inpainting · Staircasing

Mathematics Subject Classification (2000) 49J52 · 65K10 · 65N21 · 90C25 · 52A41 · 68U10

1 Introduction

We consider the following general framework of a combined first and second order non-smooth regularisation procedure. For a given datum $u_0 \in L^s(\Omega)$, $\Omega \subset \mathbb{R}^n$, $s = 1, 2$, we compute a regularised reconstruction u as a minimiser of a combined first and second order functional $H(u)$. More precisely, we are interested in solving

$$\min_u \Big\{ H(u) = \frac{1}{s}\int_\Omega (u_0 - Tu)^s \, dx + \alpha \int_\Omega f(\nabla u) \, dx + \beta \int_\Omega g(\nabla^2 u) \, dx \Big\}, \tag{1.1}$$

for $s \in \{1, 2\}$, non negative regularisation parameters α, β , with convex functions $f : \mathbb{R}^2 \to \mathbb{R}^+$, $g : \mathbb{R}^4 \to \mathbb{R}^+$ with at most linear growth at infinity, and a suitable linear operator T, see Section 3 for details. By combining nonlinear first and second order dynamics in the regulariser this approach is able to balance the preservation of jumps (taken care of by the first order term in the regulariser) against the smoothness within homogeneous image parts (modeled by the linearly increasing penalisation of the Hessian $\nabla^2 u$).

We prove the well-posedness of (1.1) via the method of relaxation and discuss its relation to some higher order approaches in the literature. Its numerical solution is presented for the case $f(x) = |x|$, $g(x) = |x|$ and $f(x) = |x|_1$, $g(x) = |x|_1$ using a splitting technique and its application for image denoising, deblurring and inpainting is discussed.

1.1 Context

A general inverse problem in imaging reads as follows. Observing or measuring data $u_0 \in \mathcal{H}$ in a suitable Hilbert space of real functions defined on a domain Ω, we seek for the original or reconstructed image u that fulfils the model

$$u_0 = Tu, \tag{1.2}$$

where T is a linear operator in $\mathcal{H}$, i.e., $T \in \mathcal{L}(\mathcal{H})$. The operator T is the forward operator of this problem. Examples are blurring operators (in which case Tu denotes the convolution of u with a blurring kernel), $T = \mathcal{F}$ the Fourier transform or $T = \mathcal{P}_n$ a projection

K. Papafitsoros
Cambridge Center for Analysis, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA,
E-mail: k.papafitsoros@maths.cam.ac.uk

C.B.Schönlieb
Cambridge Center for Analysis, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA,
E-mail: C.B.Schoenlieb@damtp.cam.ac.uk

operator onto a subspace of $\mathcal{H}$ (i.e., only a few samples of u are given).

To reconstruct u one has to invert the operator T. This is not always possible since in many applications this inversion is ill-posed and further complicated by interferences like noise. In this case a common procedure in inverse problems is to add a-priori information to the model, which in general is given by a certain regularity assumption on the image u. Hence, instead of solving (1.2) one computes u as a minimiser of

$$\mathcal{J}(u) = D\left(u_0, Tu\right) + \alpha\psi(u),$$

defined in a suitable Banach space $\mathcal{H}_\psi$. Here, ψ models the regularity assumption on u with a certain regularity parameter $\alpha > 0$ and is called the regulariser of the problem, and D is a distance function in $\mathcal{H}$. The latter depends on the statistics of the data u_0, which can be either estimated or are known from the physics behind the acquisition of u_0. For normally distributed u_0, i.e., the interferences in u_0 are Gaussian noise, this distance function is the squared L^2 norm of $u_0 - Tu$. For the choice of the regulariser ψ squared Hilbert space norms have a long tradition in inverse problems. The most prominent example is H^1 regularisation

$$\min_{u\in H^1}\left\{\frac{1}{2}\|u_0 - Tu\|^2_{L^2(\Omega)} + \alpha\|\nabla u\|^2_{L^2(\Omega)}\right\}, \tag{1.3}$$

see also [53,46]. For $T = Id$ this procedure is equivalent to Gaussian filtering of the given image u_0, where the variance of the Gaussian kernel is related to the choice of α. The result of such a regularisation technique is a linearly, i.e., isotropically, smoothed image u, for which the smoothing strength does not depend on u_0. Hence, while eliminating the disruptive noise in the given data u_0 also prominent structures like edges in the reconstructed image are blurred. This observation gave way to a new class of non-smooth norm regularisers, which aim to eliminate noise and smooth the image in homogeneous areas in the image, while preserving the relevant structures such as object boundaries and edges. More precisely, instead of (1.3) one considers the following functional over the space $W^{1,1}(\Omega)$:

$$\mathcal{J}(u) = \frac{1}{2}\|u_0 - Tu\|^2_{L^2(\Omega)} + \int_\Omega f(\nabla u)\, dx, \tag{1.4}$$

where f is a function from $\mathbb{R}$ to $\mathbb{R}^+$ with at most linear growth, see [48]. As stated in (1.4) the minimisation of $\mathcal{J}$ over $W^{1,1}(\Omega)$ is not well-posed in general. For this reason relaxation procedures are applied, which embed the optimisation for $\mathcal{J}$ into the optimisation for its lower semicontinuous envelope within the larger space of functions of bounded variation, see Section 2. The most famous example in image processing is $f(s) = |s|$, which for $T = Id$ results in the so-called ROF model [41]. In this case the relaxed formulation of (1.4) is the total variation denoising model, where $\|\nabla u\|_{L^1(\Omega)}$ is replaced by the total variation $|Du|(\Omega)$ and $\mathcal{J}$ is minimised over the space of functions of bounded variation. Other examples for f are regularised versions of the total variation like $f(s) = \sqrt{s^2 + \epsilon^2}$ for a positive $\epsilon \ll 1$ [1,28] and alike [17,32,13,37]. The consideration of such regularised versions of $|\nabla u|$ is sometimes necessary for the numerical solution of (1.4) e.g., by means of time-stepping [48] or multigrid-methods [49,30,47].

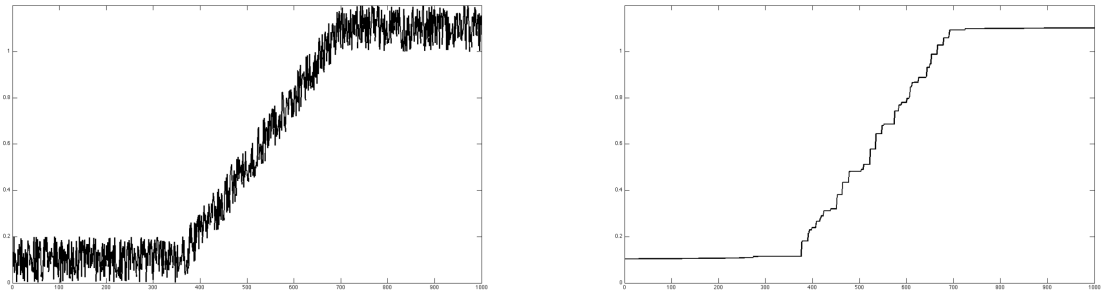
As these and many more contributions in the image processing community have proven, this new non-smooth regularisation procedure indeed results into a nonlinear smoothing of the image, smoothing more in homogeneous areas of the image domain and preserving characteristic structures such as edges. In particular, since the total variation regulariser is tuned towards the preservation of edges, this image model is performing very well if the reconstructed image is piecewise constant. The drawback of such a regularisation procedure becomes apparent as soon as images or signals (in 1D) are considered which not only consist of flat regions and jumps, but also possess slanted regions, i.e., piecewise linear parts. The effect of total variation regularisation in this case is so-called staircasing. Roughly this means that the total variation regularisation of a noisy linear function u_0 in one dimension is a staircase u, whose L^2 norm is close to u_0, see Figure 1. In one dimension this effect has been rigorously studied in [20]. In two dimensions this effect results in blocky-like images, see Figure 2.

This undesirable feature of the total variation regularisers gave reason to proposing image enhancement methods which involve higher order derivatives in the regulariser. Especially in recent years, higher order versions of non-smooth image enhancing methods have been considered.

In this paper we introduce the general higher order model (1.1), which both subsumes many existing higher order approaches (see sections 1.2-1.3) and through its generality gives rise to new models in this context. As such, its mathematical and numerical discussion is novel and now presented in this paper. We prove existence and (under certain conditions) uniqueness of minimisers of (1.1) by the method of relaxation and present its numerical solution for image denoising, deblurring and inpainting by a splitting technique. In particular, we discuss the application of (1.1) to the removal of additive and impulse noise as well as image deblurring and inpainting problems.

1.2 Related work

Already in the pioneering paper of Chambolle and Lions [13] the authors propose a higher order method by means of an inf-convolution of two convex regularisers. Here, a noisy image is decomposed into three



(a) Noisy, piecewise linear signal

(b) First order total variation denoised signal

Fig. 1 Illustration of the staircasing effect in one space dimension

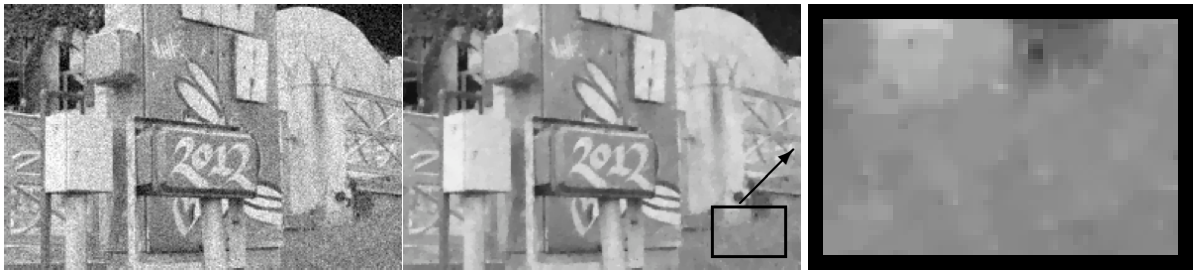


Fig. 2 First order image denoising and the staircasing effect: (l.) noisy image, (m.) denoised image, (r.) detail of the bottom right hand corner of the denoised image to visualise the staircasing effect (the creation of blocky-like patterns due to the first order regulariser)

parts $u_0 = u_1 + u_2 + n$ by solving

$$\min_{(u_1,u_2)} \Big\{ \frac{1}{2}\|u_0 - u_1 - u_2\|^2_{L^2(\Omega)} + \alpha \int_\Omega |\nabla u_1| \, dx + \beta \int_\Omega |\nabla^2 u_2| \, dx \Big\}, \tag{1.5}$$

where $\nabla^2 u_2$ is the distributional Hessian of u_2. Then, u_1 is the piecewise constant part of u_0, u_2 the piecewise smooth part and n the noise (or texture). Along these lines a modified infimal-convolution approach has been recently proposed in the discrete setting in [44,45]. Another attempt to combine first and second regularisation originates from Chan, Marquina, and Mulet [15], who consider total variation minimisation together with weighted versions of the Laplacian. More precisely, they consider a regularising term of the form

$$\alpha \int_\Omega |\nabla u| \, dx + \beta \int_\Omega f(|\nabla u|)(\Delta u)^2 \, dx,$$

where f must be a function with certain growth conditions at infinity in order to allow jumps. The well-posedness of the latter in one space dimension has been rigorously analysed by Dal Maso, Fonseca, Leoni and Morini [20] via the method of relaxation.

The idea of bounded Hessian regularisers was also considered by Lysaker et al. [35,36], Chan et al. [16], Scherzer et al. [42,33], Lai at al. [34], and Bergounioux and Piffet [5]. In these works the considered model has the general form

$$\min_u \left\{ \frac{1}{2}\|u_0 - u\|^2_{L^2(\Omega)} + \alpha |\nabla^2 u|(\Omega) \right\}.$$

Further, in [40] minimisers of functionals which are regularised by the total variation of the $(l-1)$st derivative, i.e.,

$$|D\nabla^{l-1}u|(\Omega),$$

are studied. Another interesting higher order total variation model is proposed by Bredies et al. [8]. The considered regulariser is called total generalised variation (TGV) and is of the form

$$TGV^k_\alpha(u) = \sup \Big\{ \int_\Omega u \mathrm{div}^k \xi \, dx : \xi \in C^k_c(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)), \|\mathrm{div}^l \xi\|_\infty \le \alpha_l, \; l = 0, \dots, k-1 \Big\}, \tag{1.6}$$

where $\mathrm{Sym}^k(\mathbb{R}^d)$ denotes the space of symmetric tensors of order k with arguments in $\mathbb{R}^d$, and α_l are fixed positive parameters. Its formulation for the solution of general inverse problems was given in [9].

Properties of higher order regularisers in terms of diffusion filters are further studied in [23]. Therein, the authors consider the Euler-Lagrange equations corresponding to minimisers of functionals of the general type

$$\mathcal{J}(u) = \int_\Omega (u_0 - u)^2 \, dx + \alpha \int_\Omega f\left(\sum_{|\beta|=p} |D^\beta u|^2 \right) \, dx,$$

for different non-quadratic penaliser functions f. Moreover, Bertozzi and Greer [6] have rigorously studied the fourth-order evolution equation which arises as a gradient flow of $\int G(\Delta u)$, where G is a nondecreasing

function of quadratic growth in a neighbourhood of 0 and at most linear growth at infinity. Solutions of this model are called low curvature image simplifiers and are given by

$$u_t = -\alpha\Delta(\arctan(\Delta u)) + (u_0 - u).$$

when $G(s) = s\arctan(s) - 1/2\log(s^2+1)$.

1.3 Relation of our model to TGV, infimal convolution regularisation and higher order diffusion filters

In this section we want to analyse the connection of our combined 1st and 2nd order approach (1.1) with infimal convolution (1.5) [13,44,45] and the generalised total variation regulariser (1.6) of order two [8].

In the case of inf-convolution (1.5) the regularised image $u = u_1 + u_2$ consists of a function $u_1 \in BV(\Omega)$ and a function $u_2 \in BH(\Omega)$ which are balanced against each other by positive parameters α, β. Differently, a minimiser u of (1.1) is in $BH(\Omega)$ as a whole and as such is more regular than the infimal convolution minimiser which is a function in $BV(\Omega)$. Hence, infimal convolution reproduced edges in an image as perfect jumps while in our combined total variation - bounded Hessian approach edges are lines where the image function has a large but finite gradient everywhere. Our approach (1.1) can be made equivalent to infimal convolution if combined with the correct choice of adaptive regularisation, e.g., [24,29]. More precisely, we replace the two constant parameters α and β by spatially varying functions $\alpha(x), \beta(x)$ and minimise for u

$$\frac{1}{2}\int_\Omega (u_0 - u)^2\, dx + \int_\Omega \alpha|\nabla u|\, dx + \int_\Omega \beta|\nabla^2 u|\, dx.$$

Then, we can choose α and β according to (1.5), i.e., $\alpha = 0$ where $u = u_2$, $\beta = 0$ where $u = u_1$, and α/β correctly balancing u_1 and u_2 in the rest of Ω.

The relation of (1.1) to the regularisation approach with total generalised variation [8] of order 2 can be understood through its equivalence with the modified infimal convolution approach [45] in the discrete setting. The total generalised variation of order 2 is defined for a positive multi-index $\alpha = (\alpha_0, \alpha_1)$ as

$$\begin{aligned} TGV^2_\alpha(u) = \sup\Big\{ \int_\Omega u\ \mathrm{div}^2 v\ dx : \\ v \in C^2_c(\Omega, \mathrm{Sym}^2(\mathbb{R}^d)), \\ \|\mathrm{div}^l v\|_{L^\infty(\Omega)} \le \alpha_l,\ l = 0,1\Big\}, \end{aligned}$$

where $\mathrm{Sym}^2(\mathbb{R}^d)$ is the space of symmetric tensors of order 2 with arguments in $\mathbb{R}^d$. For smooth enough functions u the second order TGV can be written as

$$\begin{aligned} TGV^2_\alpha(u) = \inf_{\substack{u_1+u_2=u \\ w_1 \in \ker\epsilon^1,\ w_2 \in \ker\epsilon^2}} \Big\{\alpha \int_\Omega |\nabla^2 u_1 + w_1|\, dx \\ +\beta \int_\Omega |\nabla u_2 + w_2|\, dx\Big\}, \end{aligned}$$

for a positive multiindex (α, β) and $\epsilon^l = \frac{1}{2}(\nabla u_l + \nabla u_l^\perp)$. As such TGV^2_α is equivalent to the modified infimal convolution regulariser proposed in [45].

Finally, the relation to higher order diffusion filters as analysed in [23] becomes apparent when considering the Euler-Lagrange equation of (1.1) in the case $T = Id$ and f, g have the form $f(x) = h(|x|)$, $g(x) = h(|x|)$, where h is convex and has at most linear growth. Namely, with appropriate boundary conditions we obtain the following Euler-Lagrange equation

$$\begin{aligned} u - u_0 = \alpha\,\mathrm{div}\left(h'(|\nabla u|)\frac{\nabla u}{|\nabla u|}\right) \\ -\beta\mathrm{div}^2\left(h'(|\nabla^2 u|)\frac{\nabla^2 u}{|\nabla^2 u|}\right). \end{aligned}$$

This simplifies for the case $h(s) = \sqrt{|s|^2 + \epsilon^2}$ to a regularised total variation/bounded Hessian flow that reads

$$\begin{aligned} u - u_0 = \alpha\,\mathrm{div}\left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \epsilon^2}}\right) \\ -\beta\mathrm{div}^2\left(\frac{\nabla^2 u}{\sqrt{|\nabla^2 u|^2 + \epsilon^2}}\right). \end{aligned}$$

The consideration of corresponding evolution equations for (1.7) in the presence of different choices of penalising functions h promises to give rise to additional properties of this regularisation technique and is a matter of future research.

For more discussion and comparison of higher order regularisers we recommend Chapters 4.1.5-4.1.7 and 6.4 in [3].

2 Preliminaries

Remarks on standard notation: As usual, we denote with $\mathcal{L}^n$ and $\mathcal{H}^n$ the Lebesgue measure and the n-dimensional Hausdorff measure on $\mathbb{R}^n$. Different notions are denoted by $|\cdot|$: When it is applied on vectors or matrices it denotes the Euclidean norm (vector) or the Frobenious norm (matrices). When it is applied on measures it denotes the total variation measure while when it is applied on Borel subsets of $\mathbb{R}^n$ it denotes the Lebesgue measure of that subset. Finally, $|\cdot|_1$ denotes the ℓ_1 norm in $\mathbb{R}^n$ and $(\cdot,\cdot)$ denotes the standard Euclidean inner product.

2.1 Finite Radon measures

All our notation and definitions are consistent with [2]. From now on, Ω will denote an open set in $\mathbb{R}^n$. We define the space $[\mathcal{M}(\Omega)]^m$ to be the space of $\mathbb{R}^m$-valued finite Radon measures. The total variation measure of $\mu \in [\mathcal{M}(\Omega)]^m$ is denoted by $|\mu|$. We say that a sequence $(\mu_k)_{k\in\mathbb{N}}$ in $[\mathcal{M}(\Omega)]^m$ converges weakly* to a

measure $\mu \in [\mathcal{M}(\Omega)]^m$ if $\lim_{k\to\infty} \int_\Omega u\, d\mu_k = \int_\Omega u\, d\mu$ for all $u \in C_0(\Omega)$ which is the completion of $C_c(\Omega)$ endowed with the supremum norm. We will often consider the Lebesgue decomposition of a $\mathbb{R}^m$-valued finite Radon measure ν with respect to a σ-finite positive Borel measure μ :

$$\nu = \nu^a + \nu^s = \left(\frac{\nu}{\mu}\right)\mu + \nu^s$$

where ν^a is the absolutely continuous part of ν with respect to μ, ν^s is the singular part and the (ν/μ) denotes the density function of ν with respect to μ (Radon-Nikodým derivative). Recall also that any $\mu \in [\mathcal{M}(\Omega)]^m$ is absolutely continuous with respect to $|\mu|$ and thus we get the *polar decomposition* of μ

$$\mu = \left(\frac{\mu}{|\mu|}\right)|\mu|, \quad with \quad \left|\frac{\mu}{|\mu|}\right| = 1, \ \ |\mu| \ \ a.e..$$

2.2 Convex functions of measures

Let g be a continuous function from $\mathbb{R}^m$ to $\mathbb{R}$ which is positively homogeneous of degree 1, i.e., for every $x \in \mathbb{R}^m$

$$g(tx) = tg(x), \quad \forall t \geq 0.$$

Given a measure $\mu \in [\mathcal{M}(\Omega)]^m$, we define the $\mathbb{R}$-valued measure $g(\mu)$ as follows:

$$g(\mu) := g\left(\frac{\mu}{|\mu|}\right)|\mu|.$$

It can be proved that if g is a convex function then $g(\cdot)$ is a convex function in $[\mathcal{M}(\Omega)]^m$ and if ν is any positive measure such that μ is absolutely continuous with respect to ν then

$$g(\mu) = g\left(\frac{\mu}{\nu}\right)\nu.$$

We refer the reader to Proposition A.1 in Appendix A for a proof of the above statement. Suppose now more generally that g is a continuous function from $\mathbb{R}^m \to \mathbb{R}$ which is convex and has at most linear growth at infinity, i.e., there exists a positive constant K such that

$$|g(x)| \leq K(1 + |x|), \quad \forall x \in \mathbb{R}^m.$$

In that case the *recession function* g_∞ of g is well defined everywhere where

$$g_\infty(x) := \lim_{t\to\infty} \frac{g(tx) - g(0)}{t}, \quad \forall x \in \mathbb{R}^m.$$

It can be proved that g_∞ is a convex function and positively homogeneous of degree 1. Given a measure $\mu \in [\mathcal{M}(\Omega)]^m$ we consider the Lebesgue decomposition with respect to the Lebesgue measure $\mathcal{L}^n$, $\mu = (\mu/\mathcal{L}^n)\mathcal{L}^n + \mu^s$ and we define the $\mathbb{R}$-valued measure $g(\mu)$ as follows:

$$\begin{aligned} g(\mu) &= g\left(\frac{\mu}{\mathcal{L}^n}\right)\mathcal{L}^n + g_\infty(\mu^s) \\ &= g\left(\frac{\mu}{\mathcal{L}^n}\right)\mathcal{L}^n + g_\infty\left(\frac{\mu^s}{|\mu^s|}\right)|\mu^s|. \end{aligned} \tag{2.1}$$

We refer the reader to Theorem A.2 in Appendix A for a result regarding lower semicontinuity of convex functions of measures with respect to the weak* convergence.

2.3 The space $[BV(\Omega)]^m$

We recall that a function $u \in L^1(\Omega)$ is said to be a function of bounded variation or else $u \in BV(\Omega)$ if its distributional derivative can be represented by a $\mathbb{R}^n$-valued finite Radon measure. This means that

$$\int_\Omega \partial_i\phi\, dx = -\int_\Omega \phi\, dD_i u, \quad \forall \phi \in C_c^1(\Omega),\ i = 1, \ldots, n.$$

for some R^n-valued finite Radon measure $Du = (D_1u, \ldots, D_nu)$. It is immediate that $W^{1,1}(\Omega) \subseteq BV(\Omega)$ since if $u \in W^{1,1}(\Omega)$ then $Du = \nabla u \mathcal{L}^n$. Consistently, we say that a function $u = (u^1, \ldots, u^m) \in [L^1(\Omega)]^m$ belongs to $[BV(\Omega)]^m$ if

$$\int_\Omega u^a \partial_i\phi\ dx = -\int_\Omega \phi\ dD_i u^a, \quad i = 1, \ldots, n,$$

$a = 1, \ldots, m$.

In that case Du is an $m \times n$ matrix valued measure. A function u belongs to $[BV(\Omega)]^m$ if and only if its variation in Ω, $V(u, \Omega)$ is finite, where,

$$V(u,\Omega) = \sup\left\{\sum_{a=1}^m \int_\Omega u^a \mathrm{div}\phi^a\ dx : \right.$$
$$\left. \phi \in [C_c^1(\Omega)]^{mn},\ \|\phi\|_\infty \leq 1\right\}.$$

Moreover if $u \in [BV(\Omega)]^m$ then $|Du|(\Omega) = V(u, \Omega)$ and if $u \in [W^{1,1}(\Omega)]^m$, then $|Du|(\Omega) = \int_\Omega |\nabla u| dx$, where $|\nabla u| = \left(\sum_{a=1}^m \sum_{i=1}^n (\partial_i u^a)^2\right)^{1/2}$. The space $[BV(\Omega)]^m$ endowed with the norm $\|u\|_{BV(\Omega)} := \int_\Omega |u| dx + |Du|(\Omega)$ is a Banach space. It can be shown that if $Du = 0$, then u is equal to a constant a.e. in any connected component of Ω.

Suppose that $(u_k)_{k\in\mathbb{N}}$, u belong to $[BV(\Omega)]^m$. We say that the sequence $(u_k)_{k\in\mathbb{N}}$ converges to u *weakly** in $[BV(\Omega)]^m$ if it converges to u in $[L^1(\Omega)]^m$ and the sequence of measures Du_k converge weakly* to the measure Du. It is known that $(u_k)_{k\in\mathbb{N}}$ converges to u weakly* in $[BV(\Omega)]^m$ if and only if $(u_k)_{k\in\mathbb{N}}$ is bounded in $[BV(\Omega)]^m$ and converges to u in $[L^1(\Omega)]^m$. The usefulness of the introduction of the weak* convergence is revealed in the following compactness result: Suppose that the sequence $(u_k)_k \in \mathbb{N}$ is bounded in $[BV(\Omega)]^m$, where Ω is a bounded open set of $\mathbb{R}^n$ with Lipschitz boundary. Then there exists a subsequence $(u_{k_\ell})_{\ell\in\mathbb{N}}$

and a function $u \in [BV(\Omega)]^m$ such that $(u_{k_\ell})_{\ell\in\mathbb{N}}$ converges to u weakly* in $[BV(\Omega)]^m$.

We say that the sequence $(u_k)_{k\in\mathbb{N}}$ converges to u *strictly* in $[BV(\Omega)]^m$ if it converges to u in $[L^1(\Omega)]^m$ and $|Du_k|(\Omega)$ converges to $|Du|(\Omega)$. It is immediate that strict convergence implies weak* convergence.

Suppose that Ω is a bounded open subset of $\mathbb{R}^n$ with Lipschitz boundary. Let $1^* = n/(n-1)$ when $n > 1$ and $1^* = \infty$ when $n = 1$. Then $BV(\Omega) \subseteq L^{1^*}(\Omega)$ with continuous embedding. Moreover if Ω is connected then the following inequality holds (Poincaré-Wirtinger):

$$\|u - u_\Omega\|_{L^{1^*}(\Omega)} \leq C|Du|(\Omega), \quad \forall u \in BV(\Omega),$$

where the constant C depends only on Ω and u_Ω denotes the mean value of u in Ω:

$$u_\Omega := \frac{1}{|\Omega|}\int_\Omega u\ dx.$$

We refer the reader to Appendix B for an introduction to weak continuity and differentiabilty notions in $BV(\Omega)$ as well as the decomposition of the distributional derivative of a function $u \in BV(\Omega)$.

2.4 Relaxed functionals

Suppose that X is a set endowed with some topology τ and let $F : X \to \mathbb{R} \cup \{+\infty\}$. The *relaxed functional* or otherwise called the *lower semicontinuous envelope* of F with respect to the topology τ is a functional $\overline{F} : X \to \mathbb{R}\cup\{+\infty\}$ defined as follows for every $x \in X$:

$$\begin{aligned}\overline{F}(x) = \ &\sup\{G(x) :\ G : X \to \mathbb{R} \cup \{+\infty\},\\ &\tau\text{-lower semicontinuous, } G(y) \leq F(y),\\ &\forall y \in X\}.\end{aligned}$$

It is easy to check that $\overline{F}$ is the greatest τ-lower semicontinuous functional which is smaller or equal than F. It can be also checked that

$$\overline{F}(x) = \sup_{U\in\mathcal{N}(x)} \inf_{y\in U} F(y),$$

where $\mathcal{N}(x)$ denotes the open neighbourhoods of x. Moreover, if X is a first countable topological space, then $\overline{F}(x)$ is characterised by the two following properties:

(i) For every sequence $(x_k)_{k\in\mathbb{N}}$ converging to x, we have

$$\overline{F}(x) \leq \liminf_{k\to\infty} F(x_k).$$

(ii) There exists a sequence $(x_k)_{k\in\mathbb{N}}$ converging to x such that

$$\overline{F}(x) \geq \limsup_{k\to\infty} F(x_k).$$

An interesting property of the relaxed functional is that if it has a minimum point then the value of $\overline{F}$ at that point will be equal with the infimum of F i.e.

$$\min_{x\in X} \overline{F}(x) = \inf_{x\in X} F(x).$$

For more information on relaxed functionals see [19] and [7].

3 The variational formulation

In the following Ω denotes as usual a bounded, connected, open subset of $\mathbb{R}^2$ with Lipschitz boundary, T denotes a bounded linear operator from $L^2(\Omega)$ to $L^2(\Omega)$, $u_0 \in L^2(\Omega)$ and α, β are non negative constants. Further we suppose that $f : \mathbb{R}^2 \to \mathbb{R}^+$, $g : \mathbb{R}^4 \to \mathbb{R}^+$ are convex functions with at most linear growth at infinity, i.e., there exist positive constants K_1 and K_2 such that

$$f(x) \leq K_1(1 + |x|), \quad \forall x \in \mathbb{R}^2, \tag{3.1}$$
$$g(x) \leq K_2(1 + |x|), \quad \forall x \in \mathbb{R}^4. \tag{3.2}$$

Moreover, we assume that both f and g are minimised in 0 and they satisfy a coercivity condition:

$$f(x) \geq K_3|x|, \quad \forall x \in \mathbb{R}^2, \tag{3.3}$$
$$g(x) \geq K_4|x|, \quad \forall x \in \mathbb{R}^4, \tag{3.4}$$

where K_3 and K_4 are strictly positive. We want to minimise the following functional:

$$\begin{aligned}H(u) = \frac{1}{2}\int_\Omega (u_0 - T(u))^2\ dx + \alpha\int_\Omega f(\nabla u)\ dx\\ +\beta\int_\Omega g(\nabla^2 u)\ dx.\end{aligned} \tag{3.5}$$

The natural space for the functional H to be defined in is $W^{2,1}(\Omega)$. Since this space is not reflexive a solution of the minimisation problem by the direct method of calculus of variations does not work. Rather, existence of a minimiser of (3.5) can be shown via relaxation that is: : We extend the functional H into a larger space which has some useful compactness properties with respect to some topology and we identify the relaxed functional with respect to the same topology. This space will be $BH(\Omega)$.

3.1 The space $BH(\Omega)$

The space $BH(\Omega)$ (often denoted with $BV^2(\Omega)$) is the space of *functions of bounded Hessian*. It was introduced by Demengel in [21] and consists of all functions $u \in W^{1,1}(\Omega)$ whose distributional Hessian can be represented by an $R^2 \times R^2$-valued finite Radon measure. In other words:

$$BH(\Omega) = \{u \in W^{1,1}(\Omega) :\ \nabla u \in [BV(\Omega)]^2\}.$$

We set $D^2u := D(\nabla u)$. Again it is immediate that $W^{2,1}(\Omega) \subseteq BH(\Omega)$. $BH(\Omega)$ is a Banach space equipped

with the norm $\|u\|_{BH(\Omega)} = \|u\|_{BV(\Omega)} + |D^2u|(\Omega)$. If Ω has a Lipschitz boundary and it is connected then it can be shown that there exist positive constants C_1 and C_2 such that

$$\int_\Omega |\nabla u(x)|\ dx \le C_1|D^2u|(\Omega) + C_2\int_\Omega |u(x)|\ dx, \quad \forall u \in BH(\Omega).$$

Moreover, the embedding from $BH(\Omega)$ into $W^{1,1}(\Omega)$ is compact, see [21]. We denote the approximate differential of ∇u with $\nabla^2 u$, see Appendix B for the definition.

Analogously with $BV(\Omega)$ we define the following notions of convergence in $BH(\Omega)$:

Definition 3.1 (Weak* convergence in BH(Ω)). *Let $(u_k)_{k\in\mathbb{N}}$, u belong to $BH(\Omega)$. We will say that u_k converges to u weakly* in $BH(\Omega)$ if*

$$u_k \to u, \quad \text{in } L^1(\Omega)$$

and

$$\nabla u_k \rightharpoonup \nabla u \quad \text{weakly}^* \text{ in } [BV(\Omega)]^2, \quad \text{as } k\to\infty,$$

or in other words

$$\|u_k - u\|_{L^1(\Omega)} \to 0,$$

$$\|\nabla u_k - \nabla u\|_{[L^1(\Omega)]^2} \to 0,$$

$$\int_\Omega \phi\ dD^2u_k \to \int_\Omega \phi\ dD^2u,\ \ \forall \phi \in C_0(\Omega).$$

It is not hard to check that a basis for that topology consists of the following sets:

$$U(v,F,\epsilon) = \Bigg\{ u \in BH(\Omega) :\ \|v-u\|_{L^1(\Omega)} + \|\nabla v - \nabla u\|_{[L^1(\Omega)]^2} + \left|\int_\Omega \phi_i\ dD^2v - \int_\Omega \phi_i\ dD^2u\right| < \epsilon, \ i \in F \Bigg\},$$

where $v \in BH(\Omega)$, F is finite, $\epsilon > 0$ and $\phi_i \in C_0(\Omega)$. This topology is also metrizable, hence first countable. We don not imply here that $BH(\Omega)$ is the dual space of a Banach space but we name this convergence *weak** to show the correspondence with the weak* convergence in $BV(\Omega)$. We have also the corresponding compactness result:

Theorem 3.2 (Compactness in BH(Ω)). *Suppose that the sequence $(u_k)_{k\in\mathbb{N}}$ is bounded in $BH(\Omega)$. Then there exists a subsequence $(u_{k_\ell})_{\ell\in\mathbb{N}}$ and a function $u \in BH(\Omega)$ such that u_{k_ℓ} converges to u weakly* in $BH(\Omega)$.*

Proof. From the compact embedding of $BH(\Omega)$ into $W^{1,1}(\Omega)$ and the fact that the sequence $(\nabla u_k)_{k\in\mathbb{N}}$ is bounded in $[BV(\Omega)]^2$ we have that there exists a subsequence $(u_{k_\ell})_{\ell\in\mathbb{N}}$, a function $u \in W^{1,1}(\Omega)$ and a function $v \in [BV(\Omega)]^2$ such that u_{k_ℓ} converges to u in $W^{1,1}(\Omega)$ and ∇u_{k_ℓ} converges to v weakly* in $[BV(\Omega)]^2$, as ℓ goes to infinity. Then, $\nabla u = v$, $u \in BH(\Omega)$ and u_{k_ℓ} converges to u weakly* in $BH(\Omega)$. □

Definition 3.3 (Strict convergence in BH). *Let $(u_k)_{k\in\mathbb{N}}$, u belong to $BH(\Omega)$. We say that u_k converges to u strictly in $BH(\Omega)$ if*

$$u_k \to u, \quad \text{in } L^1(\Omega)$$

and

$$|D^2u_k|(\Omega) \to |D^2u|(\Omega), \quad \text{as } k\to\infty.$$

It is easily checked that the function

$$d(u,v) = \int_\Omega |u-v|\ dx + \left||D^2u|(\Omega) - |D^2v|(\Omega)\right|,$$

is a metric and induces the strict convergence in $BH(\Omega)$. The following Lemma can be used to compare these two topologies.

Lemma 3.4. *Suppose that $(u_k)_{k\in\mathbb{N}}$, u belong to $BH(\Omega)$ and u_k converges to u strictly in $BH(\Omega)$. Then*

$$\|u_k - u\|_{W^{1,1}(\Omega)} \to 0, \quad \text{as } k\to\infty.$$

Proof. We recall from (3.6) that there exist positive constants C_1 and C_2 such that

$$\int_\Omega |\nabla u|\ dx \le C_1|D^2u|(\Omega) + C_2\int_\Omega |u|\ dx, \quad \forall u \in BH(\Omega).$$

Since the sequence $(u_k)_{k\in\mathbb{N}}$ is strictly convergent in $BH(\Omega)$, the sequences $(\|u_k\|_{L^1(\Omega)})_{k\in\mathbb{N}}$ and $(|D^2u_k|(\Omega))_{k\in\mathbb{N}}$ are bounded. Hence, there exists a positive constant C such that

$$\int_\Omega |\nabla u_k|\ dx < C, \quad \forall k \in \mathbb{N},$$

which implies that the sequence $(u_k)_{k\in\mathbb{N}}$ is bounded in $BH(\Omega)$. From the compact embedding of $BH(\Omega)$ into $W^{1,1}(\Omega)$ we get that there exists a subsequence $(u_{k_\ell})_{\ell\in\mathbb{N}}$ and a function $v \in W^{1,1}(\Omega)$ such that u_{k_ℓ} converges to v in $W^{1,1}(\Omega)$. In particular u_{k_ℓ} converges to v in $L^1(\Omega)$ so $v = u$ and thus u_{k_ℓ} converges to u in $W^{1,1}(\Omega)$. Since every subsequence of $(u_k)_{k\in\mathbb{N}}$ is bounded in $BH(\Omega)$ we can repeat the same argument and deduce that for every subsequence of $(u_k)_{k\in\mathbb{N}}$ there exists a further subsequence which converges to u in $W^{1,1}(\Omega)$. This proves that the initial sequence $(u_k)_{k\in\mathbb{N}}$ converges to u in $W^{1,1}(\Omega)$ □

Corollary 3.5. *Strict convergence implies weak* convergence in $BH(\Omega)$.*

3.2 Relaxation of the second order functional

We now extend the functional H to $BH(\Omega)$ in the following way:

$$H_{ex}(u) = \begin{cases} \frac{1}{2}\int_\Omega (u_0 - Tu)^2 \, dx \\ +\alpha \int_\Omega f(\nabla u) \, dx \\ +\beta \int_\Omega g(\nabla^2 u) \, dx & \text{if } u \in W^{2,1}(\Omega), \\ +\infty & \text{if } f \in \\ & BH(\Omega) \setminus W^{2,1}(\Omega). \end{cases} \tag{3.6}$$

Indeed, as we have discussed above the weak* topology in $BH(\Omega)$ provides with a good compactness property which is inherited from the weak* topology in $[BV(\Omega)]^n$. However, the functional H_{ex} is not sequentially lower semicontinuous with respect to the strict topology in $BH(\Omega)$ and hence it is neither with respect to the weak* topology in $BH(\Omega)$. Indeed, we can find a function $u \in BH(\Omega) \setminus W^{2,1}(\Omega)$, see [21] for such an example. Hence, from the definition of H_{ex} we have $H_{ex}(u) = \infty$. However, according to Theorem A.3 we can find a sequence $(u_k)_{k\in\mathbb{N}}$ in $W^{2,1}(\Omega)$ which converges strictly to u. It follows that the sequences $(\|u_k\|_{L^1(\Omega)})_{k\in\mathbb{N}}$, $(|D^2 u_k|(\Omega))_{k\in\mathbb{N}}$ as well as $(\|\nabla u_k\|_{L^1(\Omega)})_{k\in\mathbb{N}}$ are bounded. Moreover the sequence $(u_k)_{k\in\mathbb{N}}$ will be bounded in $L^2(\Omega)$. Since T is a bounded linear operator and from the fact that f and g have at most linear growth at infinity we deduce that the sequence $H_{ex}(u_k)$ is bounded as well. Hence we get

$$H_{ex}(\Omega) > \liminf_{k\to\infty} H_{ex}(u_k),$$

which proves that H_{ex} is not lower semicontinuous with respect to the strict topology in $BH(\Omega)$. Thus, we have to identify its lower semicontinuous envelope with respect to the strict convergence. This is called the method of relaxation. We define the following functional in $BH(\Omega)$:

$$\begin{aligned} \overline{H}_{ex}(u) &:= \frac{1}{2}\int_\Omega (u_0 - Tu)^2 dx + \alpha \int_\Omega f(\nabla u) \, dx \\ &\quad +\beta \, g(D^2 u)(\Omega) \\ &= \frac{1}{2}\int_\Omega (u_0 - Tu)^2 dx + \alpha \int_\Omega f(\nabla u) \, dx \\ &\quad +\beta \int_\Omega g(\nabla^2 u) \, dx \\ &\quad +\beta \int_\Omega g_\infty \left(\frac{D^s \nabla u}{|D^s \nabla u|} \right) d|D^s \nabla u|, \end{aligned}$$

where $\nabla^2 u$, the approximate differential of ∇u, is also the density of $D^2 u$ with respect to the Lebesgue measure, see Appendix B. It is immediate to see that if $u \in W^{2,1}(\Omega)$ then $\overline{H}_{ex}(u) = H_{ex}(u)$. Thus is general, $\overline{H}_{ex}$ is smaller than H_{ex}.

Theorem 3.6. *The functional $\overline{H_{ex}}$ is lower semicontinuous with respect to the strict topology in $BH(\Omega)$.*

Proof. It is not hard to check that since f is convex and it has at most linear growth then it is Lipschitz, say with constant $L > 0$. Let u and $(u_k)_{k\in\mathbb{N}}$ be functions in $BH(\Omega)$ and let u_k converge to u strictly in $BH(\Omega)$ and thus also weakly* in $BH(\Omega)$. We have to show that

$$\overline{H_{ex}}(u) \le \liminf_{k\to\infty} \overline{H_{ex}}(u_n).$$

From the definition of the weak* convergence in $BH(\Omega)$ we have that u_k converges to u in $W^{1,1}(\Omega)$. From the Sobolev inequality, see [27],

$$\|v\|_{L^2(\Omega)} \le C\|v\|_{W^{1,1}(\Omega)}, \quad \forall v \in W^{1,1}(\Omega),$$

we have that u_k converges to u in $L^2(\Omega)$. Since $T : L^2(\Omega) \to L^2(\Omega)$ is continuous then the map $u \mapsto \frac{1}{2}\int_\Omega (u_0 - Tu)^2 dx$ is continuous and hence we have that

$$\frac{1}{2}\int_\Omega (u_0 - Tu_k)^2 dx \to \frac{1}{2}\int_\Omega (u_0 - Tu)^2 dx, \quad \text{as } k \to \infty. \tag{3.7}$$

Moreover since $\|\nabla u_k - \nabla u\|_{[L^1(\Omega)]^2}$ converges to 0 as $k \to \infty$, we have from the Lipschitz property

$$\left| \int_\Omega f(\nabla u_k) \, dx - \int_\Omega f(\nabla u) \, dx \right| \le$$

$$\int_\Omega |f(\nabla u_k) - f(\nabla u)| \, dx \le$$

$$L \int_\Omega |\nabla u_k - \nabla u| \, dx \to 0, \quad \text{as } k \to \infty,$$

i.e., we have

$$\int_\Omega f(\nabla u_k) \, dx \to \int_\Omega f(\nabla u) \, dx, \quad \text{as } k \to \infty. \tag{3.8}$$

Finally we have that $D^2 u_k \to D^2 u$ weakly*. We can apply Theorem A.2 for $\mu_k = \mu = \mathcal{L}^2$, $\nu = D^2 u$, $\nu_k = D^2 u_k$ and get that

$$g(D^2 u)(\Omega) \le \liminf_{k\to\infty} g(D^2 u_k)(\Omega). \tag{3.9}$$

From (3.7), (3.8) and (3.9) we get that

$$\overline{H_{ex}}(u) \le \liminf_{k\to\infty} \overline{H_{ex}}(u_k).$$

□

Theorem 3.7. *The functional $\overline{H_{ex}}$ is the lower semicontinuous envelope of H_{ex} with respect to the strict convergence in $BH(\Omega)$.*

Proof. Suppose that $(u_k)_{k\in\mathbb{N}}$ converges to u strictly in $BH(\Omega)$. From the lower semicontinuity of $\overline{H_{ex}}$ we have that $\overline{H_{ex}}(u) \le \liminf_{k\to\infty} \overline{H_{ex}}(u_k)$ and since $\overline{H_{ex}} \le H_{ex}$ we get that $\overline{H_{ex}}(u) \le \liminf_{k\to\infty} H_{ex}(u_k)$. For the other direction Theorem A.3 tells us that given $u \in BH(\Omega)$ there exist a sequence $(u_k)_{k\in\mathbb{N}} \subseteq C^\infty(\Omega) \cap$

$W^{2,1}(\Omega)$ such that u_k converges to u strictly in $BH(\Omega)$ and

$$g(D^2u_k)(\Omega) \to g(D^2u)(\Omega). \tag{3.10}$$

We have also that u_k converges to u in $W^{1,1}(\Omega)$ which implies that

$$\begin{aligned}&\frac{1}{2}\int_\Omega (u_0 - Tu_k)^2 dx + \alpha \int_\Omega f(\nabla u_k)\ dx \to \\ &\frac{1}{2}\int_\Omega (u_0 - T(u))^2 dx + \alpha \int_\Omega f(\nabla u)\ dx, \\ &\text{as } k \to \infty,\end{aligned} \tag{3.11}$$

as the proof of Theorem 3.6 shows. From (3.10) and (3.11) we get that

$$\overline{H_{ex}}(u) = \lim_{k\to\infty} H_{ex}(u_k).$$

□

Remark 3.8. *Observe that $\overline{H_{ex}}$ is also the lower semicontinuous envelope of H_{ex} with respect to the weak* convergence in $BH(\Omega)$.*

The proof of the following minimisation theorem follows the proof of the corresponding theorem in [48] for the analogue first order functional.

Theorem 3.9. *Assuming $T(\mathcal{X}_\Omega) \neq 0$, $\alpha > 0$, $\beta > 0$ then the minimisation problem*

$$\inf_{u\in BH(\Omega)} \overline{H_{ex}}(u), \tag{3.12}$$

has a solution $u \in BH(\Omega)$.

Proof. Let $(u_k)_{k\in\mathbb{N}}$ be a minimising sequence for (3.12) and let $C > 0$ be an upper bound for $(\overline{H_{ex}}(u_k))_{k\in\mathbb{N}}$. We have that

$$\int_\Omega f(\nabla u_k)\ dx < C \quad \text{and} \quad \frac{1}{2}\int_\Omega (u_0 - Tu_k)^2 dx < C, \tag{3.13}$$

for every $k \in \mathbb{N}$ From the coercivity assumptions (3.3)-(3.4) and from (3.13) we have

$$|Du_k|(\Omega) = \int_\Omega |\nabla u_k|\ dx < C, \quad \forall k \in \mathbb{N}, \tag{3.14}$$

for a possibly different constant C. We show that the sequence $(u_k)_{k\in\mathbb{N}}$ is bounded in $L^2(\Omega)$, following essentially [48]. By the Poincaré-Wirtinger inequality there exists a positive constant C_1 such that for every $k \in \mathbb{N}$

$$\begin{aligned}\|u_k\|_{L^2(\Omega)} &= \left\| u_k - \mathcal{X}_\Omega \frac{1}{|\Omega|}\int u_k\ dx \right. \\ &\qquad \left. + \mathcal{X}_\Omega \frac{1}{|\Omega|}\int u_k\ dx \right\|_{L^2(\Omega)} \\ &\leq C_1 |Du_k|(\Omega) + \left| \int_\Omega u_k\ dx \right| \\ &\leq C + \left| \int_\Omega u_k\ dx \right|.\end{aligned}$$

Thus it suffices to bound $\left|\int_\Omega u_k dx\right|$ uniformly in k. We have for every $k \in \mathbb{N}$

$$\left\| T\left(\mathcal{X}_\Omega \frac{1}{|\Omega|}\int u_k\ dx \right) \right\|_{L^2(\Omega)} \leq$$

$$\left\| T\left(\mathcal{X}_\Omega \frac{1}{|\Omega|}\int u_k\ dx \right) - Tu_k \right\|_{L^2(\Omega)}$$

$$+ \|Tu_k - u_0\|_{L^2(\Omega)} + \|u_0\|_{L^2(\Omega)} \leq$$

$$\|T\| \left\| u_k - \mathcal{X}_\Omega \frac{1}{|\Omega|}\int u_k\ dx \right\|_{L^2(\Omega)} + \|Tu_k - u_0\|_{L^2(\Omega)}$$

$$+ \|u_0\|_{L^2(\Omega)} \leq$$

$$C_1\|T\||Du_k|(\Omega) + \sqrt{2C} + \|u_0\|_{L^2(\Omega)} \leq$$

$$C_1\|T\|C + \sqrt{2C} + \|u_0\|_{L^2(\Omega)} := C'.$$

It follows that

$$\left| \int_\Omega u_n\ dx \right| \|T(\mathcal{X}_\Omega)\|_{L^2(\Omega)} \leq C'|\Omega|,$$

and thus

$$\left| \int_\Omega u_k\ dx \right| \leq \frac{C'|\Omega|}{\|T(\mathcal{X}_\Omega)\|_{L^2(\Omega)}},$$

since $T(\mathcal{X}(\Omega)) \neq 0$.

Since the sequence is bounded in $L^2(\Omega)$ and Ω is bounded, we have that the sequence is bounded in $L^1(\Omega)$ and moreover it is bounded in $BH(\Omega)$. From Theorem 3.2 we obtain the existence of a subsequence $(u_{k_\ell})_{\ell\in\mathbb{N}}$ and $u \in BH(\Omega)$ such that u_{k_ℓ} converges to u weakly* in $BH(\Omega)$. Since the functional $\overline{H}_{ex}$ is lower semicontinuous with respect to this convergence we have:

$$\overline{H_{ex}}(u) \leq \liminf_{k\to\infty} \overline{H_{ex}}(u_n)$$

which implies that

$$u = \min_{u\in BH(\Omega)} \overline{H_{ex}}(u).$$

□

Let us note here that in the above proof we needed $\alpha > 0$, in order to get an à priori bound in the L^1 norm of the gradient (for $\beta = 0$ see [48]). However, the proof goes through if $\alpha = 0$ and T is injective. If T is not injective and $\alpha = 0$ it is not straightforward how to get existence. The proof of the following theorem also follows the proof of the corresponding theorem for the first order analogue in [48].

Proposition 3.10. *If T is injective or if f is strictly convex, then the solution of the minimisation problem (3.12) is unique.*

Proof. Using the Proposition A.1 in Appendix A we can check that the functional $\overline{H_{ex}}$ is convex. Let u_1, u_2 be two minimisers. If $T(u_1) \neq T(u_2)$ then from the strict convexity of the first term of $\overline{H_{ex}}$ we have

$$\overline{H_{ex}}\left(\frac{1}{2}u_1 + \frac{1}{2}u_2\right) < \frac{1}{2}\overline{H_{ex}}(u_1) + \frac{1}{2}\overline{H_{ex}}(u_2) =$$

$$\inf_{u\in BH(\Omega)} \overline{H_{ex}}(u)$$

which is impossible. Hence $T(u_1) = T(u_2)$ and if T is injective we have $u_1 = u_2$. If T is not injective but f is strictly convex then we must have $\nabla u_1 = \nabla u_2$ otherwise we get the same contradiction as before. Since Ω is connected there exists a constant c such that $u_1 = u_2 + c\mathcal{X}_\Omega$ and since $T(\mathcal{X}_\Omega) \neq 0$, we get $c = 0$. □

4 Special cases and extensions

In this section we introduce two more versions of the functional $\overline{H_{ex}}$, the anisotropic case and the case where the L^1 norm appears in the fidelity term. As we will see in the numerical experiments these models are more effective in specific cases.

4.1 The anisotropic case

We introduce the anisotropic version of the functional H in (3.5). Note that when $f(x) = |x|$, $g(x) = |x|$ then the relaxed functional $\overline{H}_{ex}$ is given by

$$\frac{1}{2}\int_\Omega (u_0 - Tu)^2 dx + \alpha \int_\Omega |\nabla u| \; dx + \beta |D^2 u| \Omega,$$

Its anisotropic analogue is defined for $f(x) = |x|_1$ and $g(x) = |x|_1$. In that case, the relaxed functional is given by

$$\begin{aligned} F(u) = \frac{1}{2}\int_\Omega (u_0 - Tu)^2 \; dx + \alpha \int_\Omega (|u_x| + |u_y|) \; dx \\ +\beta\big(|D_1 u_x|(\Omega) + |D_2 u_x|(\Omega) \\ +|D_1 u_y|(\Omega) + |D_2 u_y|(\Omega)\big). \end{aligned} \tag{4.1}$$

As we will see in the numerical implementations the minimisation of the anisotropic version gives better results when the gradient of the image is paraller or vertical to the x-axis. Since the functional F is obtained for the above choice of f and g, the following theorem holds as a special case of Theorem 3.9:

Theorem 4.1. *Assuming $T(\mathcal{X}_\Omega) \neq 0$, the minimisation problem*

$$\inf_{u\in BH(\Omega)} F(u), \tag{4.2}$$

has a solution. If T is injective then the solution is unique.

4.2 L^1 fidelity term

We consider here the case with the L^1 norm in the fidelity term, i.e.

$$G(u) = \int_\Omega |u_0 - Tu| \; dx + \alpha \int_\Omega |\nabla u| \; dx + \beta |D^2 u|(\Omega). \tag{4.3}$$

where for simplicity we consider the case $f(x) = |x|$, $g(x) = |x|$. As it has been shown in [38] and also studied in [14] and [25], the L^1 norm in the fidelity term leads to efficient restorations of images that have been corrupted by impulse (speckle) noise.

Theorem 4.2. *Assuming $T(\mathcal{X}_\Omega) \neq 0$, the minimisation problem*

$$\inf_{u\in BH(\Omega)} G(u), \tag{4.4}$$

has a solution.

Proof. the proof is another application of the direct method of calculus of variation. Similarly with the proof of Theorem 3.9 we show that any minimising sequence is bounded in $L^1(\Omega)$. Hence it is bounded in $BH(\Omega)$. Thus we can extract a weakly* convergent subsequence in $BH(\Omega)$. Trivially, the functional is lower semicontinuous with respect to that convergence. □

Note that in this case the uniqueness of the minimiser cannot be guaranteed since the functional G is not strictly convex anymore, even in the case where $T = Id$. The more general case of Theorem 3.9 can be easily extended to the cases discussed in Sections 4.1 and 4.2.

5 The numerical implementation

In this section we are working with the discretised version of the functional 3.5 and we describe the numerical method we use in order to implement the minimisation, namely the split Bregman method. In particular, in Section 5.1 we introduce the discrete versions of L^1 and L^2 norms as well as the discrete gradient and Hessian. In Section 5.2 we provide an introduction to the Bregman iteration stating some of its important properties and we describe how the Bregman iteration can be used for the solution of constrained optimisation problems, an idea originated in [39]. In [31] the Bregman iteration combined with a splitting technique (split Bregman) was used in order to solve the total variation minimisation. In the latter paper it was also proved that the iterates of the Bregman iteration converge to the solution of the constrained problem assuming that the iterates satisfy the constraint in a finite number of iterations. We give a more general convergence result where we do not use that assumption, see Theorem 5.5. Finally in Section 5.3

we describe how our problem can be solved with the Bregman iteration, using the splitting procedure mentioned above.

5.1 The discretised setting

In this section we study the discretisation and minimisation of the functional 3.5. In our numerical examples we consider two different choices for f and g. The first one is isotropic regularisation with $f(x) = |x|$, $g(x) = |x|$ that is to say when

$$\overline{H_{ex}}(u) = \frac{1}{2}\int_\Omega (u_0 - Tu)^2 dx + \alpha \int_\Omega |\nabla u|\, dx + \beta |D^2 u|(\Omega). \tag{5.1}$$

The second choice is anisotropic regularisation with $f(x) = |x|_1$, $g(x) = |x|_1$. In that case the functional has the form

$$\begin{aligned}\overline{H_{ex}}(u) = \frac{1}{2}\int_\Omega (u_0 - Tu)^2 dx + \alpha \int_\Omega (|u_x| + |u_y|)\, dx \\ +\beta\big(|D_1 u_x|(\Omega) + |D_2 u_x|(\Omega) \\ +|D_1 u_y|(\Omega) + |D_2 u_y|(\Omega)\big).\end{aligned} \tag{5.2}$$

Moreover, the data fidelity term is either measured in the L^2 or in the L^1 norm. For these choices we have to define a couple of discrete operators and norms. We will denote the discretised version of (5.1) with J. In the discretised setting, u will be an element of $\mathbb{R}^{n\times m}$, T will be a bounded linear operator from $\mathbb{R}^{n\times m}$ to $\mathbb{R}^{n\times m}$. Moreover, for $f(x) = |x|$ and $g(x) = |x|$, the discrete functional J is given by

$$J(u) = \frac{1}{2}\|u_0 - Tu\|_2^2 + \alpha\|\nabla u\|_1 + \beta\|\nabla^2 u\|_1, \tag{5.3}$$

where for every $u \in \mathbb{R}^{n\times m}$, for every $v = (v_1, v_2) \in (\mathbb{R}^{n\times m})^2$ and for every $w = (w_1, w_2, w_3, w_4) \in (\mathbb{R}^{n\times m})^4$ we define

$$\|u\|_2 := \left(\sum_{i=1}^n \sum_{j=1}^m u(i,j)^2\right)^{1/2},$$

$$\|v\|_2 := \left(\sum_{i=1}^n \sum_{j=1}^m v_1(i,j)^2 + v_2(i,j)^2\right)^{1/2},$$

$$\|w\|_2 := \left(\sum_{i=1}^n \sum_{j=1}^m w_1(i,j)^2 + w_2(i,j)^2 + w_3(i,j)^2 + w_4(i,j)^2\right)^{1/2},$$

$$\|u\|_1 := \sum_{i=1}^n \sum_{j=1}^m |u(i,j)|,$$

$$\|v\|_1 := \sum_{i=1}^n \sum_{j=1}^m \left(v_1(i,j)^2 + v_2(i,j)^2\right)^{1/2},$$

$$\|w\|_1 := \sum_{i=1}^n \sum_{j=1}^m \big(w_1(i,j)^2 + w_2(i,j)^2 + w_3(i,j)^2 + w_4(i,j)^2\big)^{1/2}.$$

We follow [12] and [5] for the formulation of the discrete gradient and Hessian of $u \in \mathbb{R}^{n\times m}$. That is, we define ∇ and div consistently with the continuous setting as adjoint operators. The same for the Hessian H and its adjoint div$H$. We have that $\nabla u = ((\nabla u)_1, (\nabla u)_2)$, where

$$(\nabla u)_1(i,j) = \begin{cases} u(i,j+1) - u(i,j) & \text{if } 1 \le i \le n, \\ & 1 \le j < m, \\ 0 & \text{if } 1 \le i \le n, \\ & j = m. \end{cases}$$

$$(\nabla u)_2(i,j) = \begin{cases} u(i,i+1) - u(i,j) & \text{if } 1 \le i < n, \\ & 1 \le j \le m, \\ 0 & \text{if } i = n, \\ & 1 \le j \le m. \end{cases}$$

and $\nabla^2 u = Hu = ((Hu)_{11}, (Hu)_{12}, (Hu)_{21}, (Hu)_{22})$, where

$$(Hu)_{11}(i,j) = \begin{cases} u(i,j+1) - 2u(i,j) & \text{if } 1 \le i \le n, \\ +u(i,j-1) & 1 < j < m, \\ u(i,j+1) - u(i,j) & \text{if } 1 \le i \le n, \\ & j = 1, \\ u(i,j-1) - u(i,j) & \text{if } 1 \le i \le n, \\ & j = m. \end{cases}$$

$$(Hu)_{22}(i,j) = \begin{cases} u(i+1,j) - 2u(i,j) & \text{if } 1 < i < n, \\ +u(i-1,j) & 1 \le j \le m, \\ u(i+1,1) - u(i,j) & \text{if } i = 1, \\ & 1 \le j \le m, \\ u(i-1,j) - u(i,j) & \text{if } i = n, \\ & 1 \le j \le m. \end{cases}$$

$$(Hu)_{12}(i,j) = \begin{cases} u(i+1,j) - u(i+1,j-1) & \text{if } 1 \le i < n, \\ -u(i,j) + u(i,j-1) & 1 < j \le m, \\ 0 & \text{if } i = n, \\ & 1 \le j \le m, \\ 0 & \text{if } 1 \le i \le n, \\ & j = 1. \end{cases}$$

$$(Hu)_{21}(i,j) = \begin{cases} u(i,j+1) - u(i-1,j+1) & \text{if } 1 < i \le n, \\ -u(i,j) + u(i-1,j) & 1 \le j < m, \\ 0 & \text{if } i = 1, \\ & 1 \le j \le m, \\ 0 & \text{if } 1 \le i \le n, \\ & j = m. \end{cases}$$

The elements $(Hu)_{11}$, (Hu_{22}), $(Hu)_{12}$, $(Hu)_{21}$ denote the discretised second derivatives u_{xx}, u_{yy}, u_{xy}, u_{yx} respectively.

We also need to define the discrete divergence and second divergence operators, div and divH, that have the properties:

$$\mathrm{div} : \left(\mathbb{R}^{n\times m}\right)^2 \to \mathbb{R}^{n\times m} \quad \text{with}$$

$$-\mathrm{div}(v)\cdot u = v\cdot \nabla u, \quad \forall u \in \mathbb{R}^{n\times m},\ v \in \left(\mathbb{R}^{n\times m}\right)^2,$$

$$\mathrm{divH} : \left(\mathbb{R}^{n\times m}\right)^4 \to \mathbb{R}^{n\times m} \quad \text{with}$$

$$\mathrm{divH}(w)\cdot u = w\cdot Hu, \quad \forall u \in \mathbb{R}^{n\times m},\ w \in \left(\mathbb{R}^{n\times m}\right)^4,$$

where the dot denotes the Euclidean inner product, if we consider u, v and w as large vectors formed successively by the columns of the corresponding matrices. We refer the reader to Appendix C for the exact formulation of these divergence operators.

5.2 Constrained optimization and the Bregman iteration

Suppose we have to solve the following constrained minimisation problem:

$$\min_{u\in\mathbb{R}^d} E(u) \text{ such that } Au = b, \tag{5.4}$$

where the function E is convex and A is a linear map from $\mathbb{R}^d$ to $\mathbb{R}^\ell$. We transform the constrained minimisation problem (5.4) into an unconstrained one, introducing a parameter λ. In order to satisfy the constraint we have to let λ go to infinity:

$$\sup_{\lambda} \min_{u\in\mathbb{R}^d} E(u) + \frac{\lambda}{2}\|Au-b\|_2^2. \tag{5.5}$$

Instead of doing that we perform the Bregman iteration as it was proposed in [39] and [54]:

Bregman Iteration

$$u_{k+1} = \min_{u\in\mathbb{R}^d} E(u) + \frac{\lambda}{2}\|Au-b_k\|_2^2, \tag{5.6}$$

$$b_{k+1} = b_k + b - Au_{k+1}. \tag{5.7}$$

In this section, we give a short description of the Bregman iteration together with some known results as these were presented in [31,39,54]. We show that under certain conditions, the iterates u_k in 5.6, converge to the solution of the constrained minimisation problem 5.4. In Section 5.3 we describe how we can transform the unconstrained minimisation problem 5.3 $\min_u J(u)$ into a constrained one and then solve it with the Bregman iteration using a splitting technique.

We start with some basic definitions. Recall that if E is a convex function from $\mathbb{R}^d$ to $\mathbb{R}$ then the *subdifferential* of E evaluated at $v \in \mathbb{R}$, denoted by $\partial E[v]$, is the following set:

$$\partial E[v] := \{p \in \mathbb{R}^d :\ E(u) \geq E(v) + (p, u-v),\ \ \forall u \in \mathbb{R}^d\}.$$

The *Bregman distance* associated with the convex function E at the point v is defined as

$$D_E^p(u,v) := E(u) - E(v) - (p, u-v)$$

where p belong to the subdifferential of E evaluated at v. From the definition $D_E^p(u,v) \geq 0$ and if E is strictly convex then $D_E^p(u,v) = 0$ if and only if $u = v$. In general $D_E^p(u,v) \neq D_E^p(v,u)$ so the Bregman distance is not a metric. However, the Bregman distance does a carry a notion of distance since for every w of the form $w = \lambda u + (1-\lambda)v$, $\lambda \in [0,1]$ we have $D_E^p(w,v) \leq D_E^p(u,v)$.

For notational convenience we set $H(u) := \|Au - b\|_2^2$. The original form for Bregman iteration as it was proposed in [10] reads:

$$u_{k+1} = \arg\min_{u\in\mathbb{R}^d} D_E^{p_k}(u, u_k) + H(u), \tag{5.8}$$

$$p_{k+1} = p_k - \nabla H(u_{k+1}). \tag{5.9}$$

The following theorem shows that iterates are well defined in the sense that p_k always belongs to the subdifferential of E evaluated at some point, except perhaps for $k = 0$. It is a simple modification in the finite dimensional setting of the corresponding proposition in [39] where we refer the reader to for a proof. In particular, in [39], they work in the continuous setting and consider the case where $E(u) = |Du|(\Omega)$ and $H(u) = \frac{\lambda}{2}\|f - Ku\|$, for $f \in L^2$ and K a bounded linear operator.As we will see in Section 5.3 in our case E will be essentially equal to J. Notice that J is strictly convex when T is injective.

Theorem 5.1. *Suppose that E is a strictly convex functional from $\mathbb{R}^d$ to $\mathbb{R}$. Moreover, assume that E is coercive i.e. there exist constants $C \geq 0$ and D such that $E(u) \geq C\|u\|_2 + D$ for all $u \in \mathbb{R}^d$. Finally let $u_0 = 0$ and $p_0 = 0$. Then for every $k \in \mathbb{N}$, the solution to the minimisation problem (5.8) is unique and $p_k \in \partial E[u_k]$ for $k = 1, 2, \ldots$.*

The following auxiliary theorem states some monotonicity properties of the Bregman iteration. Again see [39] for a proof.

Theorem 5.2. *Suppose that E is a convex functional and consider the algorithm (5.8)-(5.9) with the initialisation $u_0 = 0$ and $p_0 = 0$. Suppose that the solutions u_k to the minimisation problems (5.8) exist. Then we have:*

$$H(u_k) \leq H(u_k) + D_E^{p_{k-1}}(u_k, u_{k-1}) \leq H(u_{k-1}),\quad k = 2, 3, \ldots. \tag{5.10}$$

$$D_E^{p_k}(u, u_k) + D_E^{p_{k-1}}(u_k, u_{k-1}) + H(u_k) \leq H(u) + D_E^{p_{k-1}}(u, u_{k-1}),\quad \forall u \in \mathbb{R}^d,\ \ k = 1, 2, \ldots. \tag{5.11}$$

In particular if $\tilde{u}$ is a minimiser of H we get

$$D_E^{p_k}(\tilde{u}, u_k) \leq D^{p_{k-1}}(\tilde{u}, u_{k-1}), \quad k = 2, 3, \dots.$$

The next theorem reveals more information about the convergence of the algorithm (5.8)-(5.9). The proof that we provide is a simple adjustment in the finite dimensional setting of the corresponding proof in [39]. Moreover, our version includes a further remark, inequality (5.13), that we are going to use later.

Theorem 5.3. *Suppose that E is a convex functional. Consider the algorithm (5.8)-(5.9) with initial values $u_0 = 0$ and $p_0 = 0$. Suppose that the solutions u_k to the minimisation problems (5.8) exist. Then for a minimiser $\tilde{u}$ of H and for every $k = 2, 3, \dots$ we have:*

$$H(u_k) \leq H(\tilde{u}) + \frac{E(\tilde{u}) - E(u_1)}{k-1}, \tag{5.12}$$

and in fact

$$\sum_{k=1}^{\infty} (H(u_k) - H(\tilde{u})) < \infty. \tag{5.13}$$

which tells us that $(u_k)_{k\in\mathbb{N}}$ is a minimising sequence for H. Moreover, there exists a positive constant C such that

$$E(u_k) \leq C, \quad \forall k \in \mathbb{N}. \tag{5.14}$$

Proof. By taking the sum $\sum_{\nu=1}^{k}$ in (5.11) we get

$$D_E^{p_k}(\tilde{u}, u_k) + \sum_{\nu=1}^{k} \left(D_E^{p_{\nu-1}}(u_\nu, u_{\nu-1}) + H(u_\nu) - H(\tilde{u}) \right) \leq$$
$$D_E^0(\tilde{u}, u_0) = E(\tilde{u}) - E(0), \tag{5.15}$$

and hence

$$D_E^{p_k}(\tilde{u}, u_k) + \sum_{\nu=2}^{k} \left(D_E^{p_{\nu-1}}(u_\nu, u_{\nu-1}) + H(u_\nu) - H(\tilde{u}) \right)$$
$$\leq D_E^0(\tilde{u}, u_0) = E(\tilde{u}) - E(u_1) \geq 0. \tag{5.16}$$

From the non-negativity of the term $D_E^{p_{\nu-1}}(u_\nu, u_{\nu-1})$ and the monotonicity of $H(u_\nu)$ we get

$$D_E^{p_k}(\tilde{u}, u_k) + (k-1)\left(H(u_k) - H(\tilde{u})\right) \leq E(\tilde{u}) - E(u_1) \Rightarrow$$

$$(k-1)\left(H(u_k) - H(\tilde{u})\right) \leq E(\tilde{u}) - E(u_1),$$

which proves (5.12). In fact from (5.16) we also get that

$$\sum_{\nu=2}^{k} (H(u_\nu) - H(\tilde{u})) < E(\tilde{u}) - E(u_1), \quad \forall k \in \mathbb{N}$$

thus (5.13) follows. Recall now that $H(u) = \|Au - b\|_2^2$. Then (5.12) implies for $A\tilde{u} = b$ that

$$(k-1)\|Au_k - b\|_2^2 \leq \sum_{\nu=2}^{k} \|Au_\nu - b\|_2^2 \leq E(\tilde{u}) - E(u_1).$$

From (5.15) and (5.9) and the fact that $A\tilde{u} = b$, we get for $k \geq 2$,

$$\begin{aligned}
E(\tilde{u}) &\geq \sum_{\nu=1}^{k} D_E^{p_{\nu-1}}(u_\nu, u_{\nu-1}) + E(0) \\
&= E(u_k) - \sum_{\nu=1}^{k} (p_{\nu-1}, u_\nu - u_{\nu-1}) \\
&= E(u_k) - (p_{k-1}, u_k - \tilde{u}) \\
&\quad + \sum_{\nu=1}^{k-1} (p_\nu - p_{\nu-1}, u_\nu - \tilde{u}) \\
&= E(u_k) + \sum_{\nu=1}^{k-1} (\nabla H(u_\nu), u_k - \tilde{u}) \\
&\quad - \sum_{\nu=1}^{k-1} (\nabla H(u_\nu), u_\nu - \tilde{u}) \\
&= E(u_k) + 2\sum_{\nu=1}^{k-1} (A^\top(Au_\nu - b), u_k - \tilde{u}) \\
&\quad - 2\sum_{\nu=1}^{k-1} (A^\top(Au_\nu - b), u_\nu - \tilde{u}) \\
&= E(u_k) + 2\sum_{\nu=1}^{k-1} (Au_\nu - b, Au_k - b) \\
&\quad - 2\sum_{\nu=1}^{k-1} \|Au_\nu - b\|_2^2 \\
&\geq E(u_k) - (k-1)\|Au_k - b\|_2^2 - 3\sum_{\nu=1}^{k-1} \|Au_\nu - b\|_2^2 \\
&\geq E(u_k) - 4(E(\tilde{u}) - E(u_1)) - 3H(u_1).
\end{aligned}$$

This is because

$$2\sum_{\nu=1}^{k-1} (Au_\nu - b, Au_k - b) \geq$$
$$-\sum_{\nu=1}^{k-1} \|Au_\nu - b\|_2^2 - (k-1)\|Au_k - b\|_2^2.$$

Hence, (5.14) follows. □

Remark: The above theorem tells us that the quantities $E(u_k) + H(u_k)$ are uniformly bounded in k. From now on we assume that H has the form $H(u) = (\lambda/2)\|Au - b\|_2^2$, where λ is a positive constant. Note that all the previous theorems hold. In that case the algorithm (5.8)-(5.9) now reads:

$$u_{k+1} = \arg\min_{u\in\mathbb{R}^d} E(u) - (p_k, u - u_k) + \frac{\lambda}{2}\|Au - b\|_2^2, \tag{5.17}$$

$$p_{k+1} = p_k - \lambda A^\top(Au_{k+1} - b), \tag{5.18}$$

where we have ignored the constant $E(0)$ in the objective function in (5.17). We are going to state an equivalent formulation of the above iteration that was

investigated in [54]. We have the following two versions of the algorithm first of which is the algorithm (5.17)-(5.18).

Algorithm 1:

$$u_0 = 0, \quad p_0 = 0, \tag{5.19}$$

$$u_{k+1} = \arg\min_{u\in\mathbb{R}^d} E(u) - (p_k, u - u_k) + \frac{\lambda}{2}\|Au - b\|_2^2, \tag{5.20}$$

$$p_{k+1} = p_k - \lambda A^\top(Au_{k+1} - b). \tag{5.21}$$

Algorithm 2:

$$u_0 = 0, \quad b_0 = 0, \tag{5.22}$$

$$b_{k+1} = b_k + b - Au_k, \tag{5.23}$$

$$u_{k+1} = \arg\min_{u\in\mathbb{R}^d} E(u) + \frac{\lambda}{2}\|Au - b_k\|_2^2. \tag{5.24}$$

The following theorem from [54] shows that the two formulations are equivalent.

Theorem 5.4. *Suppose that the minimisation problem (5.20) has a unique solution for every k. Then the Algorithm 1 (5.19-5.21) and the Algorithm 2 (5.22-5.24) are equivalent in the sense that the minimisation problems (5.20) and (5.24) have the same objective functions that might differ up to a constant.*

The following theorem was proved in [31] in the case where the iterates satisfy the constraint in a finite number of iterations. Here we give a more general proof where we do not use that assumption.

Theorem 5.5. *Suppose that the constrained minimisation problem (5.4) has a unique solution u^*. Moreover, suppose that the convex functional E is positive and coercive. Finally suppose that the Theorem 5.4 holds as well. Then the sequence of the iterates $(u_k)_{k\in\mathbb{N}}$ of Bregman iteration (i.e. given by Algorithm 1 or 2) converges to u^*.*

Proof. From the fact that Theorem 5.4 holds and from (5.12)-(5.13) we get that, for some $M \geq 0$

$$\|Au_k - b\|_2 \leq \frac{M}{\sqrt{k-1}}, \quad \forall k \geq 1$$
$$\text{and } \sum_{k=1}^{\infty} \|Au_k - b\|_2^2 < \infty. \tag{5.25}$$

Moreover we have ($b_0 = 0$) that

$$b_{k+1} = \sum_{\nu=1}^{k}(b - Au_\nu) \Rightarrow \|b_{k+1}\|_2 \leq \sum_{\nu=1}^{k}\|Au_\nu - b\|_2.$$

From the Theorem 5.3 and the coercivity of E we get that the sequence $\|u_k\|_2$ is bounded, say by a constant C. Thus it suffices to show that every accumulation point of that sequence is equal to u^*. Since $Au^* = b$, we have for every increasing sequence of naturals $(k_\ell)_{\ell\in\mathbb{N}}$.

$$\begin{aligned}\frac{\lambda}{2}\|Au^* - b_{k_\ell}\|_2^2 &\leq \frac{\lambda}{2}\left(\|Au^* - Au_{k_\ell}\|_2 + \|Au_{k_\ell} - b_{k_\ell}\|_2\right)^2 \\ &= \frac{\lambda}{2}\left(\|Au_{k_\ell} - b\|_2 + \|Au_{k_\ell} - b_{k_\ell}\|_2\right)^2 \\ &\leq \frac{\lambda M}{2(k_{n_\ell}-1)} \\ &\quad +\lambda\|Au_{k_\ell} - b\|_2\|Au_{k_\ell} - b_{k_\ell}\|_2 \\ &\quad +\frac{\lambda}{2}\|Au_{k_\ell} - b_{k_\ell}\|_2^2,\end{aligned}$$

and because of the fact that

$$E(u_{k_\ell}) + \frac{\lambda}{2}\|Au_{k_\ell} - b_{k_\ell}\|_2^2 \leq E(u^*) + \frac{\lambda}{2}\|Au^* - b_{k_\ell}\|_2^2,$$

we get that

$$\begin{aligned}E(u_{k_\ell}) &\leq E(u^*) + \frac{\lambda M}{2(k_\ell - 1)} \\ &\quad +\lambda\|Au_{k_\ell} - b\|_2\|Au_{k_\ell} - b_{k_\ell}\|_2 \\ &\leq E(u^*) + \frac{\lambda M}{2(k_\ell - 1)} + \frac{\lambda M}{\sqrt{k_\ell - 1}}\|Au_{k_\ell}\|_2 \\ &\quad +\lambda\|Au_{k_\ell} - b\|_2\|b_{k_\ell}\|_2 \\ &\leq E(u^*) + \frac{\lambda M}{2(k_\ell - 1)} + \frac{\lambda\|A\|C}{\sqrt{k_\ell - 1}} \\ &\quad +\lambda\|Au_{k_\ell} - b\|_2\sum_{\nu=1}^{k_\ell}\|Au_\nu - b\|_2.\end{aligned}$$

Suppose now that u_{k_ℓ} converges to some $\tilde{u}$ as k goes to infinity then $\tilde{u}$ also satisfies $\|A\tilde{u} - b\|_2 = 0$. Then if we take limits because of Kronecker's lemma, see Lemma D.1, the right hand side limit is $E(u^*)$. We have

$$E(\tilde{u}) \leq E(u^*)$$

and since u^* is the solution of the constrained minimisation problem we have $\tilde{u} = u^*$. We conclude that the whole sequence $(u_k)_{k\in\mathbb{N}}$ converges to u^*. □

5.3 Numerical solution of our minimisation problem

In this section, we explain how the Bregman iteration together with an operator splitting technique can be used to implement numerically the minimisation of our functional. The idea originates from [31] where it is called Split Bregman algorithm. Let us mention here, however, that this iterative technique is equivalent to certain instances of combinations of the augmented Lagrangian method with classical operator splitting such as Douglas-Rachford, see [43]. We also refer the reader to the preprint of Burger, Sawatzky, Brune, Benning [4] for applications of Bregman methods to higher-order regularisation models for image reconstruction.

Exemplarily, we present the resulting algorithm for the minimisation of J in (5.3), i.e., for $f(x) = |x|$, $g(x) = |x|$ and an L^2 data fidelity term. The other instances of the general higher order model (3.5) that

we consider in Sections 6, 7 and 8 are solved with the same technique, which result into very similar algorithms as the one described in the sequel.

Recall that we want to solve the following unconstrained minimisation problem:

$$\min_{u\in\mathbb{R}^{n\times m}} \frac{1}{2}\|u_0 - Tu\|_2^2 + \alpha\|\nabla u\|_1 + \beta\|\nabla^2 u\|_1 \qquad (5.26)$$

The derivation of the Split Bregman algorithm for solving (5.26) starts with the first important observation that this above minimisation problem is equivalent to the following constrained minimisation problem:

$$\min_{\substack{u\in\mathbb{R}^{n\times m}\\ v\in\left(\mathbb{R}^{n\times m}\right)^2\\ w\in\left(\mathbb{R}^{n\times m}\right)^4}} \frac{1}{2}\|u_0 - Tu\|_2^2 + \alpha\|v\|_1 + \beta\|w\|_1, \qquad (5.27)$$

such that $v = \nabla u,\ w = \nabla^2 u$.

For every $u \in \mathbb{R}^{n\times m}$ we denote by $col(u)$ the $nm-$th column vector formed successively by the columns of u. With this notation we introduce the column vector ω

$$\omega = \big(col(u), col(v_1), col(v_2), \\ col(w_{11}), col(w_{22}), col(w_{12}), col(w_{21})\big), \qquad (5.28)$$

which has total length $7mn = d$. Since the gradient and the Hessian are linear operations it is clear that the minimisation problem (5.27) can be reformulated into the following problem:

$$\min_{\omega\in\mathbb{R}^d} E(\omega) \text{ such that } A\omega = b, \qquad (5.29)$$

where $E : \mathbb{R}^d \to \mathbb{R}^+$ is convex, A is a $d\times d$ matrix and b is a vector of length d. For completeness, we describe the form of E and A. Define the following projections P_u, P_v and P_w acting on ω such that if ω has the form (5.28) then

$$P_u(\omega) = (col(u)),$$

$$P_v(\omega) = (col(v_1), col(v_2)),$$

$$P_w(\omega) = (col(w_{11}), col(w_{22}), col(w_{12}), col(w_{21})).$$

Then we set

$$E(\omega) = \frac{1}{2}\|col(u_0) - T(P_u(\omega))\|_2^2 + \alpha\|P_v(\omega)\|_1 \\ +\beta\|P_w(\omega)\|_1, \quad \forall\omega\in\mathbb{R}^d.$$

where T is seen now as a linear operator from $\mathbb{R}^{nm}$ to $\mathbb{R}^{nm}$. As far as the constraints in 5.27 are concerned, these can be written as follows:

$$A\omega = b \Longleftrightarrow \left(\begin{array}{c|c|c} 0 & 0 & 0\\ \hline -\nabla & I & 0\\ \hline -\nabla^2 & 0 & I\end{array}\right)\left(\begin{array}{c} u\\ \hline v\\ \hline w\end{array}\right) = \left(\begin{array}{c} 0\\ \hline 0\\ \hline 0\end{array}\right).$$

It follows that b is the zero vector. Recall that for constrained minimisation problems of the type (5.27) the corresponding Bregman iterative scheme will be

$$u_{k+1} = \min_{u\in\mathbb{R}^d} E(u) + \frac{\lambda}{2}\|Au - b_k\|_2^2, \qquad (5.30)$$

$$b_{k+1} = b_k + b - Au_{k+1}. \qquad (5.31)$$

It is easy to see that the iterative scheme of the type (5.30-5.31) that corresponds to the constraint minimisation problem (5.27) will be :

$$(u^{k+1}, v^{k+1}, w^{k+1}) = \operatorname*{argmin}_{u,v,w} \frac{1}{2}\|u_0 - Tu\|_2^2 + \alpha\|v\|_1 \\ +\beta\|w\|_1 + \frac{\lambda}{2}\|b_1^k + \nabla u - v\|_2^2 \\ +\frac{\lambda}{2}\|b_2^k + \nabla^2 u - w\|_2^2, \qquad (5.32)$$

$$b_1^{k+1} = b_1^k + \nabla u^{k+1} - v^{k+1}, \qquad (5.33)$$

$$b_2^{k+1} = b_2^k + \nabla^2 u^{k+1} - w^{k+1}. \qquad (5.34)$$

where $b_1^{k+1} = (b_{1,1}^{k+1}, b_{1,2}^{k+1}) \in (\mathbb{R}^{n\times m})^2$ and $b_2^{k+1} = (b_{2,11}^{k+1}, b_{2,22}^{k+1}, b_{2,12}^{k+1}, b_{2,21}^{k+1}) \in (\mathbb{R}^{n\times m})^4$.

Remark 5.6. *Notice that at least in the case where T is the identity function, the minimisation problem (5.32) has a unique solution. Moreover, the functional E is positive and coercive and the constrained minimisation problem (5.27) has a unique solution. Thus, the iterative scheme (5.30)-(5.31) is indeed equivalent to the Bregman iteration and Theorem 5.5 holds.*

Our next concern is the efficient numerical solution of the minimisation problem (5.32). We follow [31] and iteratively minimise with respect to u, v and w alternatingly:

Split Bregman for Isotropic TV − BH − L² Standard Splitting

$$b_1^0 = 0, \quad b_2^0 = 0, \qquad (5.35)$$

$$u^{k+1} = \operatorname*{argmin}_{u\in\mathbb{R}^{n\times m}} \frac{1}{2}\|u_0 - Tu\|_2^2 \\ +\frac{\lambda}{2}\|b_1^k + \nabla u - v^k\|_2^2 \\ +\frac{\lambda}{2}\|b_2^k + \nabla^2 u - w^k\|_2^2, \qquad (5.36)$$

$$v^{k+1} = \operatorname*{argmin}_{v\in(\mathbb{R}^{n\times m})^2} \alpha\|v\|_1 \\ +\frac{\lambda}{2}\|b_1^k + \nabla u^{k+1} - v\|_2^2, \qquad (5.37)$$

$$w^{k+1} = \operatorname*{argmin}_{w\in(\mathbb{R}^{n\times m})^4} \beta\|w\|_1 \\ +\frac{\lambda}{2}\|b_2^k + \nabla^2 u^{k+1} - w\|_2^2, \qquad (5.38)$$

$$b_1^{k+1} = b_1^k + \nabla u^{k+1} - v^{k+1}, \qquad (5.39)$$

$$b_2^{k+1} = b_2^k + \nabla^2 u^{k+1} - w^{k+1}. \qquad (5.40)$$

The above alternating minimisation scheme, make up the *split Bregman iteration* that was proposed in [31] to solve the total variation minimisation problem as well as problems related to compressed sensing. For convergence properties of the split Bregman iteration and also other splitting techniques we refer the reader to [43,26,18]. In [43] the split Bregman method is proved to be equivalent with the so-called Douglas-Rachford algorithm and its convergence is guaranteed.

The first of the above minimisation problems (5.36), can be solved through its optimality condition. This condition reads:

$$\begin{aligned} &T^*Tu - \lambda \mathrm{div}\left(\nabla u\right) + \lambda \mathrm{divH}\left(Hu\right) = \\ &T^*(u_0) + \lambda \mathrm{div}\left(\left(b_1^k - v^k\right)\right) \\ &-\lambda \mathrm{divH}\left(\left(b_2^k - w^k\right)\right), \end{aligned} \tag{5.41}$$

where T^* denotes the adjoint of the discrete operator T. Since all the operators that appear in (5.41) are linear, it leads to linear system of equations with nm unknowns.

The solution of the minimisation problems (5.37) and (5.38) can be obtained exactly through a generalised *shrinkage* method. It was used in both [31] and [50]. It is a simple computation to check that if $a \in \mathbb{R}^n$ then the solution to the following problem:

$$\min_{x\in\mathbb{R}^n} \|x\|_2 + \frac{\lambda}{2}\|x-a\|_2^2, \tag{5.42}$$

can be obtained through the following formula:

$$x = \mathbb{S}_{\frac{1}{\lambda}}(a) := \max\left(\|a\|_2 - \frac{1}{\lambda}, 0\right)\frac{a}{\|a\|_2}. \tag{5.43}$$

Each of the objective functionals in (5.37) and (5.38) can be written as a sum of functionals of the same type in (5.42) where $n = 2$, 4 respectively. Thus the solution to the problems (5.37) and (5.38) can be computed as follows:

$$\begin{aligned} v^{k+1}(i,j) &= \left(v_1^{k+1}(i,j), v_2^{k+1}(i,j)\right) \\ &= \mathbb{S}_{\frac{\alpha}{\lambda}}\left(b_1^k(i,j) + \nabla u^{k+1}(i,j)\right), \end{aligned} \tag{5.44}$$

$$\begin{aligned} w^{k+1}(i,j) &= \left(w_1^{k+1}(i,j), w_2^{k+1}(i,j),\right. \\ &\qquad \left. w_3^{k+1}(i,j), w_4^{k+1}(i,j)\right) \\ &= \mathbb{S}_{\frac{\alpha}{\lambda}}\left(b_2^k(i,j) + \nabla^2 u^{k+1}(i,j)\right). \end{aligned} \tag{5.45}$$

for $i = 1,\ldots,n$ and $j = 1,\ldots,m$.

Let us note here that another formulation alternative to (5.27) is possible. In particular the problem (5.26) is also equivalent with the following:

$$\begin{aligned} &\min_{\substack{u\in\mathbb{R}^{n\times m} \\ v\in\left(\mathbb{R}^{n\times m}\right)^2 \\ w\in\left(\mathbb{R}^{n\times m}\right)^4}} \frac{1}{2}\|u_0 - Tu\|_2^2 + \alpha\|v\|_1 + \beta\|w\|_1, \\ &\text{such that } \tilde{v} = \nabla u,\ \tilde{v} = v,\ w = \nabla\tilde{v}. \end{aligned} \tag{5.46}$$

Working as previously we can derive the following splitting technique for the minimisation of 5.46.

Split Bregman for Isotropic TV – BH – $\mathbf{L^2}$
Alternative Splitting

$$\begin{aligned} u^{k+1} &= \underset{u\in\mathbb{R}^{n\times m}}{\mathrm{argmin}} \ \frac{1}{2}\|u_0 - Tu\|_2^2 \\ &\quad + \frac{\lambda}{2}\|b_1^k + \nabla u - \tilde{v}^k\|_2^2 \end{aligned} \tag{5.47}$$

$$\begin{aligned} \tilde{v}_1^{k+1} &= \underset{\tilde{v}_1\in\mathbb{R}^{n\times m}}{\mathrm{argmin}} \ \frac{\lambda}{2}\|b_{1,1}^k + \nabla_1 u^{k+1} - \tilde{v}_1\|_2^2 \\ &\quad + \frac{\lambda}{2}\|b_{2,1}^k + v_1^k - \tilde{v}_1\|_2^2 \\ &\quad + \frac{\lambda}{2}\|(b_{3,11}^k, b_{3,12}^k) + \nabla\tilde{v}_1 - (w_{11}^k, w_{12}^k)\|_2^2 \end{aligned} \tag{5.48}$$

$$\begin{aligned} \tilde{v}_2^{k+1} &= \underset{\tilde{v}_2\in\mathbb{R}^{n\times m}}{\mathrm{argmin}} \ \frac{\lambda}{2}\|b_{1,2}^k + \nabla_2 u^{k+1} - \tilde{v}_2\|_2^2 \\ &\quad + \frac{\lambda}{2}\|b_{2,2}^k + v_2^k - \tilde{v}_2\|_2^2 \\ &\quad + \frac{\lambda}{2}\|(b_{3,21}^k, b_{3,22}^k) + \nabla\tilde{v}_2 - (w_{21}^k, w_{22}^k)\|_2^2 \end{aligned} \tag{5.49}$$

$$\begin{aligned} v^{k+1} &= \underset{v\in(\mathbb{R}^{n\times m})^2}{\mathrm{argmin}} \ \alpha\|v\|_1 \\ &\quad + \frac{\lambda}{2}\|b_2^k + \tilde{v}^{k+1} - v\|_2^2, \end{aligned} \tag{5.50}$$

$$\begin{aligned} w^{k+1} &= \underset{w\in(\mathbb{R}^{n\times m})^4}{\mathrm{argmin}} \ \beta\|w\|_1 \\ &\quad + \frac{\lambda}{2}\|b_3^k + \nabla\tilde{v}^{k+1} - w\|_2^2, \end{aligned} \tag{5.51}$$

$$b_1^{k+1} = b_1^k + \nabla u^{k+1} - \tilde{v}^{k+1}, \tag{5.52}$$

$$b_2^{k+1} = b_2^k + \tilde{v}^{k+1} - v^{k+1}, \tag{5.53}$$

$$b_3^{k+1} = b_3^k + \nabla\tilde{v}^{k+1} - w^{k+1}. \tag{5.54}$$

Notice that the above splitting leads to three linear system of equations for the optimality conditions of (5.47), (5.48) and (5.49) instead of one for the standard splitting. However the matrices that correspond to these systems are sparser than the one that corresponds to the optimality condition of (5.36). As Table 1 shows, the standard splitting is a bit faster but we believe that there is room for further optimization of our code in order to explore this extra sparsity.

We have performed numerical experiments in denoising, deblurring and inpainting. In all our computations we used $\lambda = 1$. Let us note here that in all of our numerical examples the range of image values is $[0,1]$ (zero for black color and one for white).

6 Applications in Denoising

In this section we discuss the application of the TV-BH approach (5.3) to image denoising. We show experiments for different levels and kinds of noise (Gaussian and impulse noise) and with various combinations of α and β. Note that for $\beta = 0$, our model corresponds

to the classical Rudin-Osher-Fatemi denoising model. The operator T in (5.3) equals the identity. Our basic non-real world test image can be seen in Figure 3.

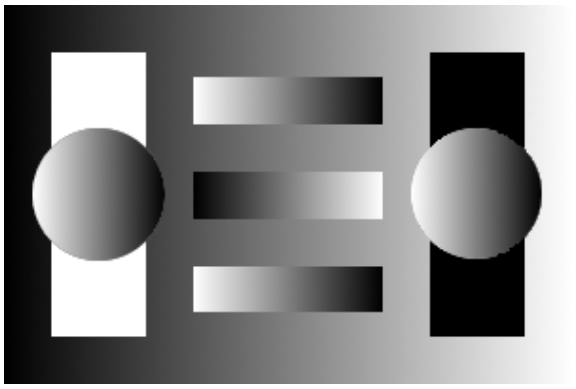


Fig. 3 Main test image. Resolution: 200× 300 pixels

Our main assessment for the quality of the reconstruction is the *structural similarity index* SSIM [52, 51]. The reason for that choice is that in contrast with traditional quality measures like the peak signal-to-noise ratio PSNR, the SSIM index also assesses the conservation of the structural information of the reconstructed image. Note that the perfect reconstruction would have SSIM value equal to 1. A justification for the choice of SSIM as a good fidelity measure instead of the traditional PSNR can be seen in Figure 4. Both images are results from denoising with the first order method ($\beta = 0$, noise: Gaussian, Variance $= 0.1$). The left picture is the one with the highest SSIM value (0.6595) while the SSIM value of the right picture is significantly lower (0.4955). This assessment comes into an agreement with the human point of view since, even though this is subjective, one would consider the left picture as a better reconstruction. On the other hand the left picture has slightly smaller PSNR value (14.02) than the right one (14.63), which was the highest PSNR value. Similar results are obtained for $\beta \neq 0$.

A stopping criterium for our algorithm we used a predefined number of iterations. In most examples this number was 300. We observed that after 80-100 iterations the relative residual of the iterates was of the order of 10^{-3} or lower (see also Table 1) and hence no noticeable change in the iterative solutions was observed after that.

In the following we shall examine wether the introduction of the higher order term ($\beta \neq 0$) in the denoising procedure, produces results of higher quality.

6.1 Isotropic denoising with L^2 fidelity term

We start with considering the minimisation of

$$J(u) = \frac{1}{2}\|u_0 - u\|_2^2 + \alpha\|\nabla u\|_1 + \beta\|\nabla^2 u\|_1,$$

which we use as an isotropic higher order TV model to eliminate Gaussian noise in the given image u_0. The noise has been produced with MATLAB's built in function *imnoise*. We compare the first order with the second order method, by showing the image that has the highest SSIM value among those that were recovered with $\beta = 0$ and $\beta \neq 0$ respectively. Note, however, that the optimal β value in terms of SSIM does not always correspond to the β value which visually reduces the staircasing effect best. The latter is in general slightly bigger than the one chosen by SSIM, see Figure 6. Still, for proof of concept, we prefer to stick with an objective quality measure and SSIM, in our opinion, is the most reliable choice for that matter.

Figure 5 depicts one denoising example, where the original image is corrupted with Gaussian noise of variance 0.005. The highest SSIM value for the isotropic TV-denoising is achieved for $\alpha = 0.12$ (SSIM=0.8979) while the highest one for the second order TV method is achieved for $\alpha = 0.06$, $\beta = 0.03$ (SSIM=0.9081). In Figure 5(b) we have plotted the middle row slices of the images (original, corrupted and restored). The reduction of the staircasing effect is clearly visible. Let us note that for larger values of β the decrease of the staircasing effect is even more noticeable but the solutions become slightly more blurry. In Figure 6 we have plotted again the image with the highest SSIM, $\alpha = 0.06$, $\beta = 0.03$ (SSIM=0.9081) and the restored image that corresponds to $\alpha = 0.06$, $\beta = 0.06$ (SSIM=0.8989). The second image is still *pleasant* to the human eye, despite being slightly more blurry, and the staircaising effect has almost disappeared.

Next we check how the SSIM and PSNR values of the restored images behave as a function of the weighting parameters α and β. In Figure 7 we plot the results for $\alpha = 0, 0.02, 0.04, \ldots, 0.3$ and $\beta = 0, 0.005, 0.01, \ldots, 0.1$. The plots suggest that both quality measures behave in a continuous way and that they have a global maximum. However, PSNR tends to rate higher those images that have been processed with a small value of β or even with $\beta = 0$ which is not the case for SSIM. An explanation for that is that higher values of β result to a further loss of contrast, see Figure 8, something that is penalised by the PSNR. The SSIM index penalises the loss of contrast as well but it also penalises the creation of the staircasing effect. This is another indication for the suitability of SSIM over PNSR. Note also, that the contrast can be recovered easily in a post-processing stage while it is not an easy task to reduce the staircasing effect using conventional processing tools.

We have also considered numerical experiments with higher noise levels. In Figure 9 the given image is corrupted with Gaussian noise of variance 0.05. In that case the highest value of SSIM for higher order denoising is achieved for $\alpha = 0.3$ and $\beta = 0.05$, SSIM=0.8226, while the highest value for TV denoising is achieved for $\alpha = 0.2$, SSIM=0.8162. Note that

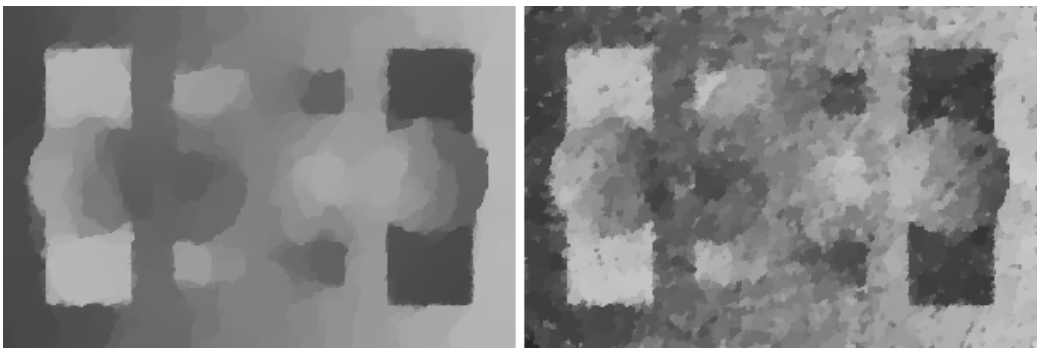


Fig. 4 Justification for the usage of SSIM index. Best SSIM value (l.) (SSIM=0.6595, PSNR=14.02) and best PSNR value (r.) (SSIM=0.4955, PSNR=14.63) among reconstructed images with the first order method ($\beta = 0$). The initial image is contaminated with Gaussian noise of variance 0.5. The better SSIM assessment of the first image agrees more with the human perception

for every fixed value of α the introduction of the higher order term is always increasing the SSIM value. However, for higher noise levels we observe in Figure 9(b) only a small improvement of the staircasing effect using the higher order denoising. Even though we see some improvement with the introduction of the second order term, it can not eliminate the artifacts in the restored image when the original image is heavily corrupted with noise.

6.2 Anisotropic denoising with L^2 fidelity term

In anisotropic denoising, our task is to implement the following minimisation problem:

$$\min_{u\in\mathbb{R}^{n\times m}} \frac{1}{2}\|u_0-u\|_2^2\alpha\left(\|(\nabla u)_1\|_1+\|(\nabla u)_2\|_1\right)$$
$$+\beta\big(\|(Hu)_{11}\|_1+\|(Hu)_{22}\|_1$$
$$+\|(Hu)_{12}\|_1+\|(Hu)_{21}\|_1\big) \tag{6.1}$$

Analogously with the isotropic case, we use a Split Bregman algorithm similar to (5.35)-(5.40) to compute a minimiser of (6.1). In short form this reads

---

Split Bregman for Anisotropic TV − BH − L² denoising

$$u^{k+1} = \operatorname*{argmin}_{u\in\mathbb{R}^{n\times m}} \frac{1}{2}\|u_0-u\|_2^2 + \frac{\lambda}{2}\|b_1^k+\nabla u-v^k\|_2^2+\frac{\lambda}{2}\|b_2^k + \nabla^2 u - w^k\|_2^2, \tag{6.2}$$

$$v^{k+1} = \operatorname*{argmin}_{v\in(\mathbb{R}^{n\times m})^2} \sum_{\ell=1}^{2}\left(\alpha\|v_\ell\|_1 \right. \tag{6.3}$$
$$\left. +\frac{\lambda}{2}\|b_{1,\ell}^k+(\nabla u^{k+1})_\ell - v_\ell\|_2^2\right), \tag{6.4}$$

$$w^{k+1} = \operatorname*{argmin}_{w\in(\mathbb{R}^{n\times m})^4} \sum_{\ell,\nu=1}^{2}\left(\beta\|w_{\ell,\nu}\|_1 \right. \tag{6.5}$$
$$\left. +\frac{\lambda}{2}\|b_{2,\ell\nu}^k+(\nabla^2 u^{k+1})_{\ell\nu}-w_{\ell\nu}\|_2^2\right), \tag{6.6}$$

$$b_1^{k+1} = b_1^k+\nabla u^{k+1}-v^{k+1}, \tag{6.7}$$
$$b_2^{k+1} = b_2^k+\nabla^2 u^{k+1}-w^{k+1}, \tag{6.8}$$

---

where $b_1^{k+1} = (b_{1,1}^{k+1}, b_{1,2}^{k+1}) \in (\mathbb{R}^{n\times m})^2$ and $b_2^{k+1} = (b_{2,11}^{k+1}, b_{2,22}^{k+1}, b_{2,12}^{k+1}, b_{2,21}^{k+1}) \in (\mathbb{R}^{n\times m})^4$. The reader should compare the algorithm (6.2)-(6.6) with the algorithm (5.36)-(5.38). The minimisation problem (6.2) is the same with (5.36) and it is solved through its optimality condition. The minimisation problems (6.4) and (6.6) can be solved exactly using shrinkage operators. Notice that the difference to the isotropic case is that v_1 and v_2 are decoupled as are w_{11}, w_{22}, w_{12}, w_{21}. Thus, if we define:

$$\mathbb{S}^{un}_{\frac{1}{\lambda}}(a) := \max\left(|a|-\frac{1}{\lambda},0\right)\frac{a}{|a|}, \quad a\in\mathbb{R},$$

then the solutions to these problems are:

$$v_\ell^{k+1}(i,j) = \mathbb{S}^{un}_{\frac{\alpha}{\lambda}}\left(b_{1,\ell}^k(i,j)+(\nabla u^{k+1})_\ell(i,j)\right),$$

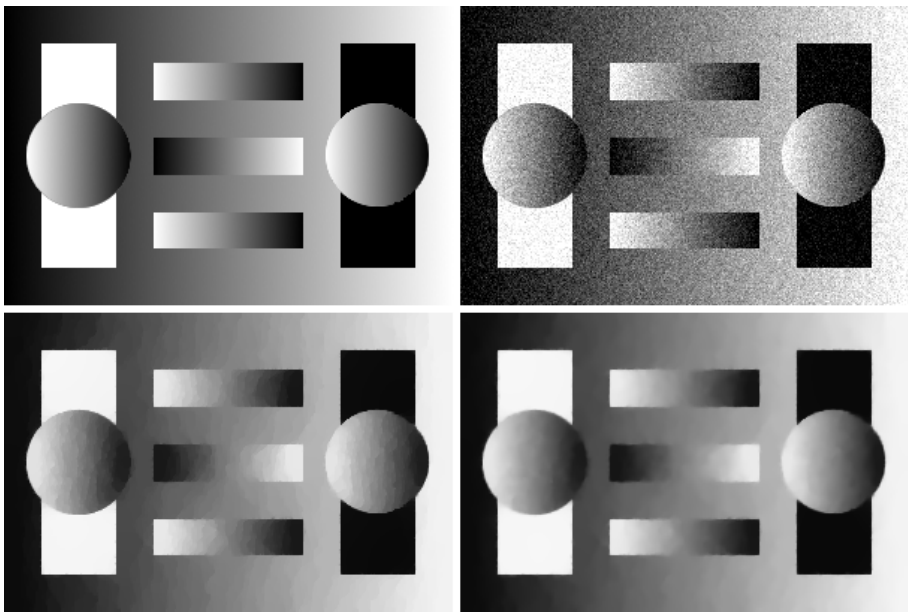
$\ell = 1,2.$

$$w_{\ell\nu}^{k+1}(i,j) = \mathbb{S}^{un}_{\frac{\alpha}{\lambda}}\left(b_{2,\ell\nu}^k(i,j)+(\nabla^2 u^{k+1})_{\ell\nu}(i,j)\right),$$

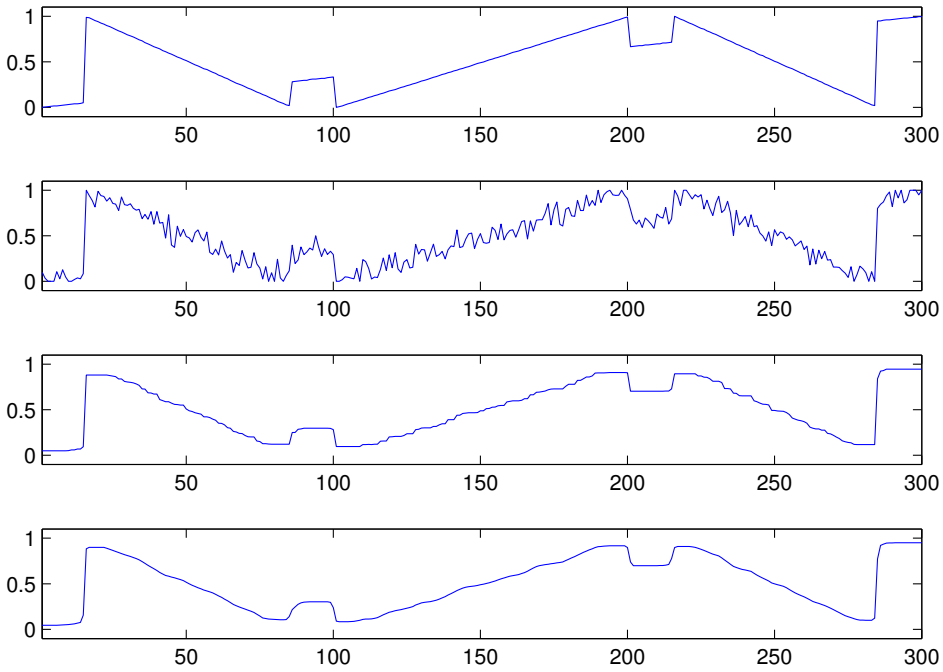
$\ell,\nu = 1,2.$

for $i = 1,\dots,n$ and $j = 1,\dots,m$.

As in the previous section we tested the algorithm for various combinations of the parameters α and β always choosing the results with optimal SSIM for presentation. In Figure 10 the best results from the first and the second order anisotropic denoising method for an image with Gaussian noise with variance = 0.005 are displayed. According to SSIM index the best TV-denoising result is the one obtained with $\alpha = 0.12$, (SSIM=0.8993). The best result for the second order denoising is obtained for $\alpha = 0.04$, $\beta = 0.025$, (SSIM=0.9164). Note that the best SSIM values of the



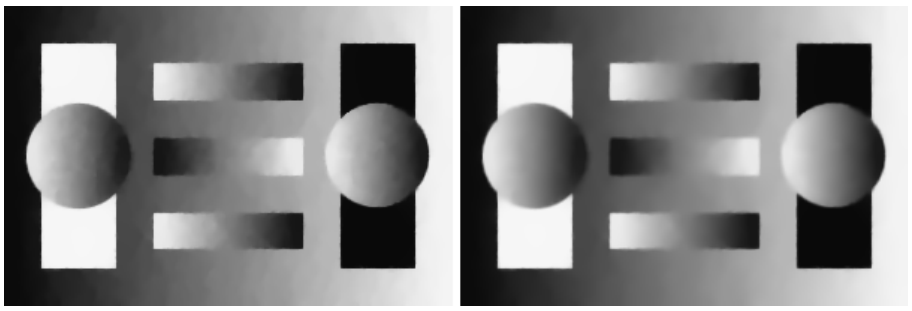
(a) (u.l.) clean image, (u.r.) image corrupted with Gaussian noise of variance 0.005 (SSIM=0.3261), (b.l) restored image with $\alpha = 0.12$, $\beta = 0$ (SSIM=0.8979), (b.r.) restored image with $\alpha = 0.06$, $\beta = 0.03$ (SSIM=0.9081)



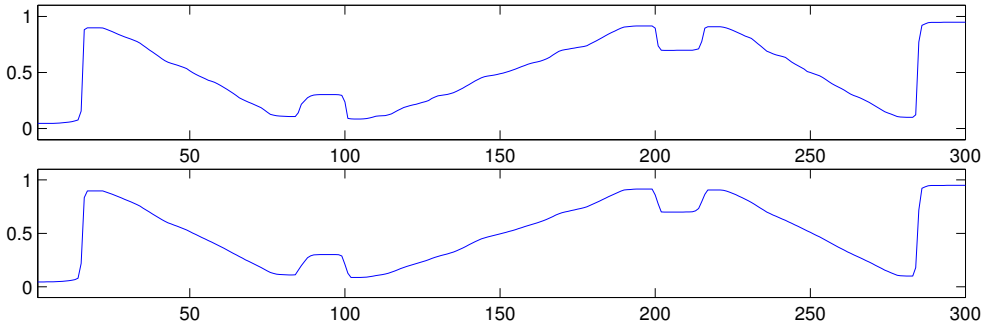
(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (u.l), (u.r), (b.l) ans (b.r.) images respectively. A decrease of the staircasing effect is noticeable

Fig. 5 Isotropic denoising. Image corrupted with Gaussian noise of variance 0.005. Number of iterations=300

images denoised anisotropicaly are slightly better than the isotropically denoised ones. The reason for this probably is the fact that in the smooth parts of our test image the direction of the gradient is parallel to the x-axis. Something that can be treated better with anisotropic denoising. For a fairer comparison we also test both methods for images whose gradient forms angles with the x-axis, see Figure 11. We observe that in the first image (first column) the anisotropic denoising performs slightly better, with SSIM values 0.8197 versus 0.8153 in the isotropic case. However in the second image (second) column where the image gradient forms various angles with the x-axis, the isotropic method is more efficient with SSIM value 0.8612 versus 0.8445 for the anisotropic.



(a) (l.) restored image with $\alpha = 0.06$, $\beta = 0.03$ (SSIM=0.9081), (r.) restored image with $\alpha = 0.06$, $\beta = 0.06$ (SSIM=0.8989)



(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (l.) and (r.) images respectively. For higher values of β we observe a decrease of the staircasing effect

Fig. 6 Isotropic denoising. Restored versions of an image corrupted with Gaussian noise of variance 0.005. Number of iterations=300. Comparison between the version with the highest SSIM value and another with the same value of α but higher value of β. The result seems more satisfactory in terms of the absence of staircasing, however higher values of β result in slightly blurrier solutions

6.3 Isotropic denoising with L^1 fidelity term

As we mentioned in Section 4.2 the L^1 norm in the fidelity term leads to more efficient denoising if the image is corrupted by impulse noise. The impulse noise is characterised by the density that counts the number of pixels that are affected by additive noise. For example, density 0.1 would mean that approximately 10% of pixels are affected. The noise is uniformly distributed on $[0, 1]$. In this case we consider the discretised version of the L^1 minimisation problem which reads:

$$\min_{u\in\mathbb{R}^{n\times m}} \|u_0 - u\|_1 + \alpha\|\nabla u\|_1 + \beta\|\nabla^2 u\|_1. \tag{6.9}$$

Similarly to the algorithms for minimising with an L^2 fidelity term, we solve the minimisation problem (6.9) using the following alternating minimisation scheme:

Split Bregman for $\mathbf{TV - BH - L^1}$ denoising

$$\begin{aligned} u^{k+1} = \underset{u}{\operatorname{argmin}}\ & \frac{\lambda}{2}\|c^k - \tilde{u}^k + u - u_0\|_2^2 \\ & + \frac{\lambda}{2}\|b_1^k + \nabla u - v^k\|_2^2 \\ & + \frac{\lambda}{2}\|b_2^k + \nabla^2 u - w^k\|_2^2, \end{aligned} \tag{6.10}$$

$$\tilde{u}^{k+1} = \underset{\tilde{u}}{\operatorname{argmin}}\ \|\tilde{u}\|_1 + \frac{\lambda}{2}\|c^k + (u^{k+1} - u_0) - \tilde{u}\|_2^2, \tag{6.11}$$

$$v^{k+1} = \underset{v}{\operatorname{argmin}}\ \alpha\|v\|_1 + \frac{\lambda}{2}\|b_1^k + \nabla u^{k+1} - v\|_2^2, \tag{6.12}$$

$$w^{k+1} = \underset{w}{\operatorname{argmin}}\ \alpha\|w\|_1 + \frac{\lambda}{2}\|b_2^k + \nabla^2 u^{k+1} - w\|_2^2, \tag{6.13}$$

$$c^{k+1} = c^k + (u^{k+1} - u_0) - \tilde{u}^{k+1}, \tag{6.14}$$

$$b_1^{k+1} = b_1^k + \nabla u_{k+1} - v^{k+1}, \tag{6.15}$$

$$b_2^{k+1} = b_2^k + \nabla^2 u_{k+1} - w^{k+1}. \tag{6.16}$$

As before the problem (6.10) is solved through its optimality condition, while the problems (6.11)-(6.13) are solved using shrinkage operators.

In Figure 12, denoising results for an image corrupted with impulse noise of density 0.4 are shown. The best results are achieved with $\alpha = 1.4$, SSIM=0.8717 for the first order method and with $\alpha = 0.9$, $\beta =$

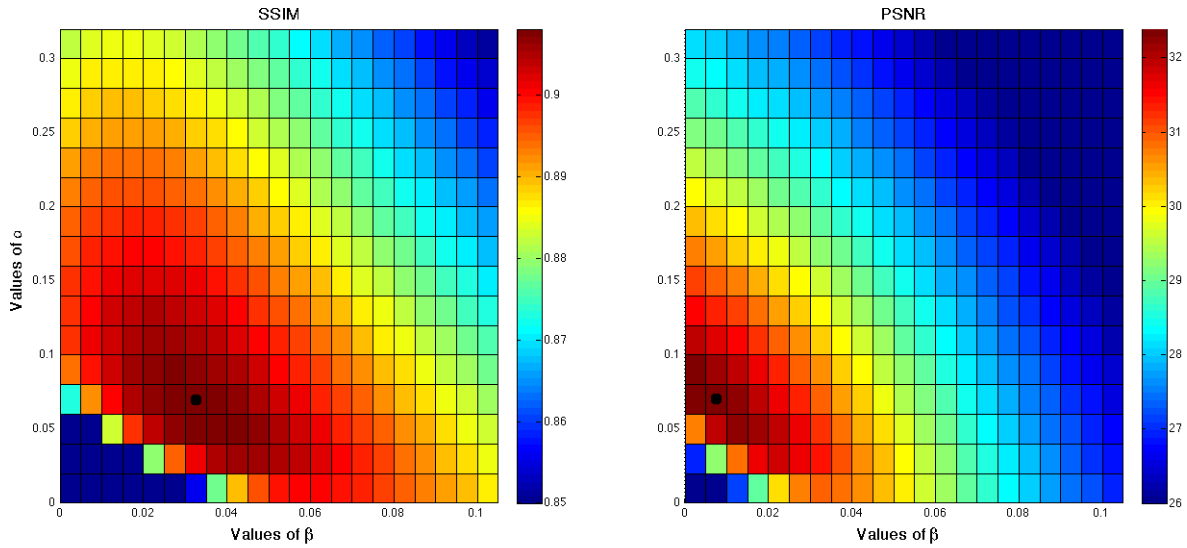


Fig. 7 Plot of the SSIM and PSNR values of the restored image as functions of α and β. For display convenience all the values under 0.85 (SSIM) and 26 (PSNR) were coloured with dark blue. The dotted cells corresponds to the highest SSIM (0.9081) and PSNR (32.39) value that were achieved for $\alpha = 0.06$, $\beta = 0.03$ and $\alpha = 0.06$, $\beta = 0.005$ respectively. Note that the first column in both plots corresponds to TV-denoising, ($\beta = 0$). The original image was corrupted with Gaussian noise of variance 0.005

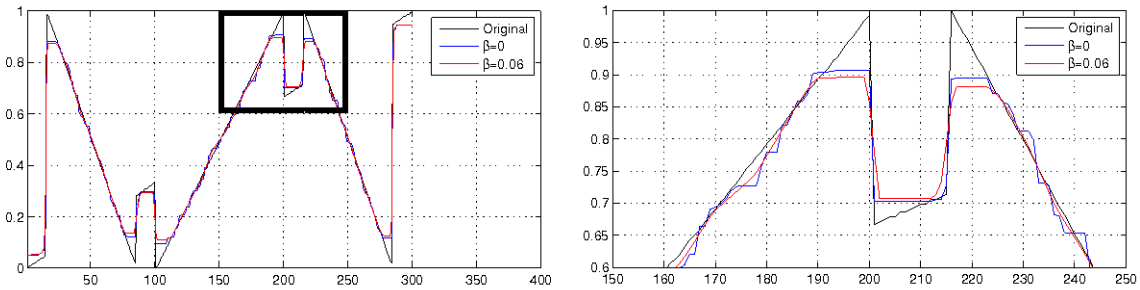


Fig. 8 Left: Middle row slices of reconstructed images with $\alpha = 0.12$, $\beta = 0$ (blue color) and $\alpha = 0.12$. $\beta = 0.06$ (red color). Slices of the original image are plotted with black color. Right: Detail of the first plot. Even though the higher order method eliminates the staircasing effect, it also results to further slight loss of contrast

0.24, SSIM=0.8817 using the second order TV-BH approach.. The almost absence of the staircasing effect in the recovered images with the first order method, does not give much space for improvement for the second order method. However, the improvement can still be observed when looking at the slices, see Figure 12(b).

7 Applications in deblurring

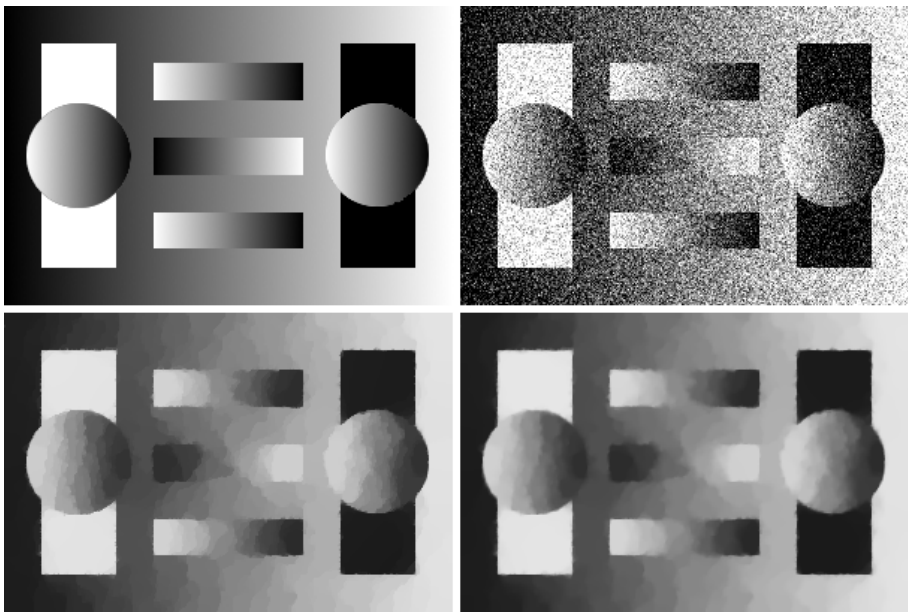
In the case of deblurring the discretised version of the minimisation problem of interest is the following:

$$\min_{u\in\mathbb{R}^{n\times m}} \frac{1}{2}\|u_0 - Tu\|_2^2 + \alpha\|\nabla u\|_1 + \beta\|\nabla^2 u\|_1, \qquad (7.1)$$

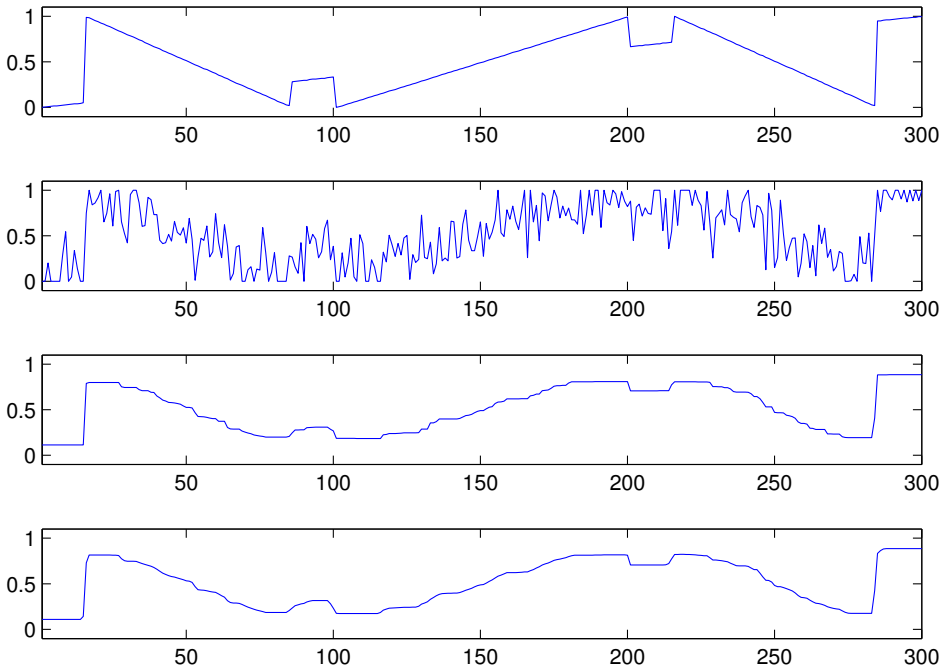
where the operator T denotes the discrete convolution with a blurring kernel. In our implementation we used discrete approximations of Gaussian kernels of variance $\sigma = 2, 4$. In order to compute the convolution, we used MATLAB's built in function *imfilter* and we defined the convolution to be circular. The minimisation procedure of the functional (7.1) was described in Section 5.3.

In Figure 13 we see the deblurring results of an image blurred after a circular convolution with a Gaussian kernel of variance $\sigma = 2$. We show the best results of the first order deblurring, $\alpha = 0.06$, SSIM=0.9066 and the overall best result, $\alpha = 0.06$, $\beta = 0.004$, SSIM=0.9173. Note that even a relatively small value of β can reduce significantly the staircasing effect, without blurring further the reconstructed image.

For higher amounts of blur, similar to high amounts of noise, the improvement is not so big anymore. In Figure 14 we consider the deblurring results for a larger blurring kernel of variance $\sigma = 4$. The best result of the first order method was obtained for $\alpha = 0.18$, SSIM=0.8105 while the overall best result was obtained for $\alpha = 0.18$, $\beta = 0.002$, SSIM=0.8129. We observe a decrease of the staircasing effect but not a total elimination. As in the denoising application a further reduction of the staircasing effect can be achieved by



(a) (u.l.) clean image, (u.r.) image corrupted with Gaussian noise of variance 0.05 (SSIM=0.1374), (b.l) restored image with $\alpha = 0.3$, $\beta = 0$ (SSIM=0.8162), (b.r.) restored image with $\alpha = 0.2$, $\beta = 0.05$ (SSIM=0.8226)



(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (u.l), (u.r), (b.l) ans (b.r.) images respectively. A decrease of the staircasing effect is visible

Fig. 9 Isotropic denoising. Image corrupted with Gaussian noise of variance 0.05. Number of iterations=300

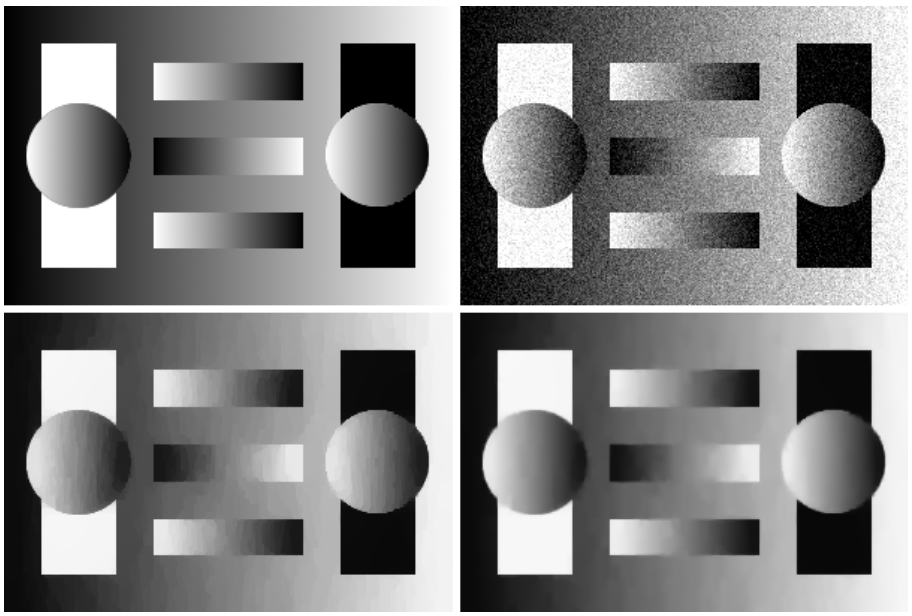
increasing the value of β, with the penalty of a slightly blurrier result, see Figure 15.

8 Applications in inpainting

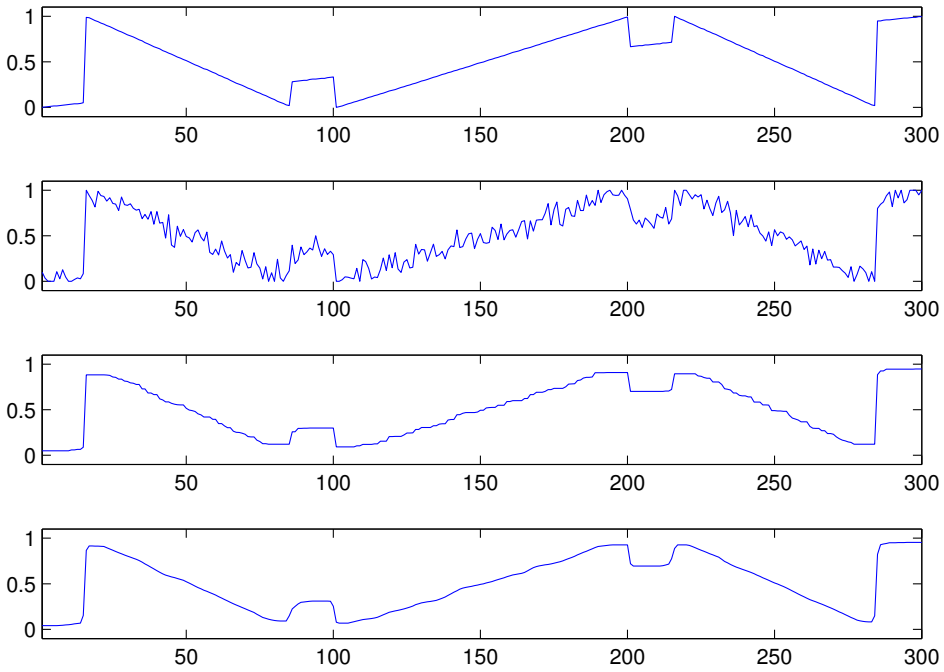
Finally, we present examples for the application of our TV-BH approach to image inpainting. There, the goal is to reconstruct an image which has a missing part using information from the intact part of the image. The missing part is a domain $D \subseteq \Omega$, known as the *inpainting domain*. In this case the operator T will be

$$Tu = \mathcal{X}_{\Omega \setminus D} u.$$

In the absence of noise one usually wants to leave the $\Omega \setminus D$ part of the image unchanged. Hence, small values of α and β are chosen. Consequently the number of iterations that are needed for the algorithm to converge is large because the heavy weighting of the fi-



(a) (u.l.) clean image, (u.r.) image corrupted with Gaussian noise of variance 0.005 (SSIM=0.3261), (b.l) restored image with $\alpha = 0.12$, $\beta = 0$ (SSIM=0.8983), (b.r.) restored image with $\alpha = 0.04$, $\beta = 0.025$ (SSIM=0.9164)



(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (u.l), (u.r), (b.l) and (b.r.) images respectively

Fig. 10 Anisotropic denoising. Image corrupted with Gaussian noise of variance 0.005. Number of iterations=200

delity term introduces a damping on the iterates of the Split Bregman iteration.

In Figure 16 the results of inpainting for two geometrical images, a stripe with a large missing part in the middle and its 45° tilted version are shown. Again we test the isotropic version versus the anisotropic version of the regularising term. Since the missing gap is large, none of the TV versions manages to connect the strip over the gap whereas in all cases the second order method manages to connect it. We notice that the anisotropic version performs slightly better in the horizontal stripe but it is unreliable in the tilted stripe. In contrast the isotropic version performs in the same manner in both cases as one would expect.

The results for inpainting a real-world image are shown in Image 17. The inpainting for the colour image has been achieved by applying the TV-BH inpainting approach to each colour channel r, g and b

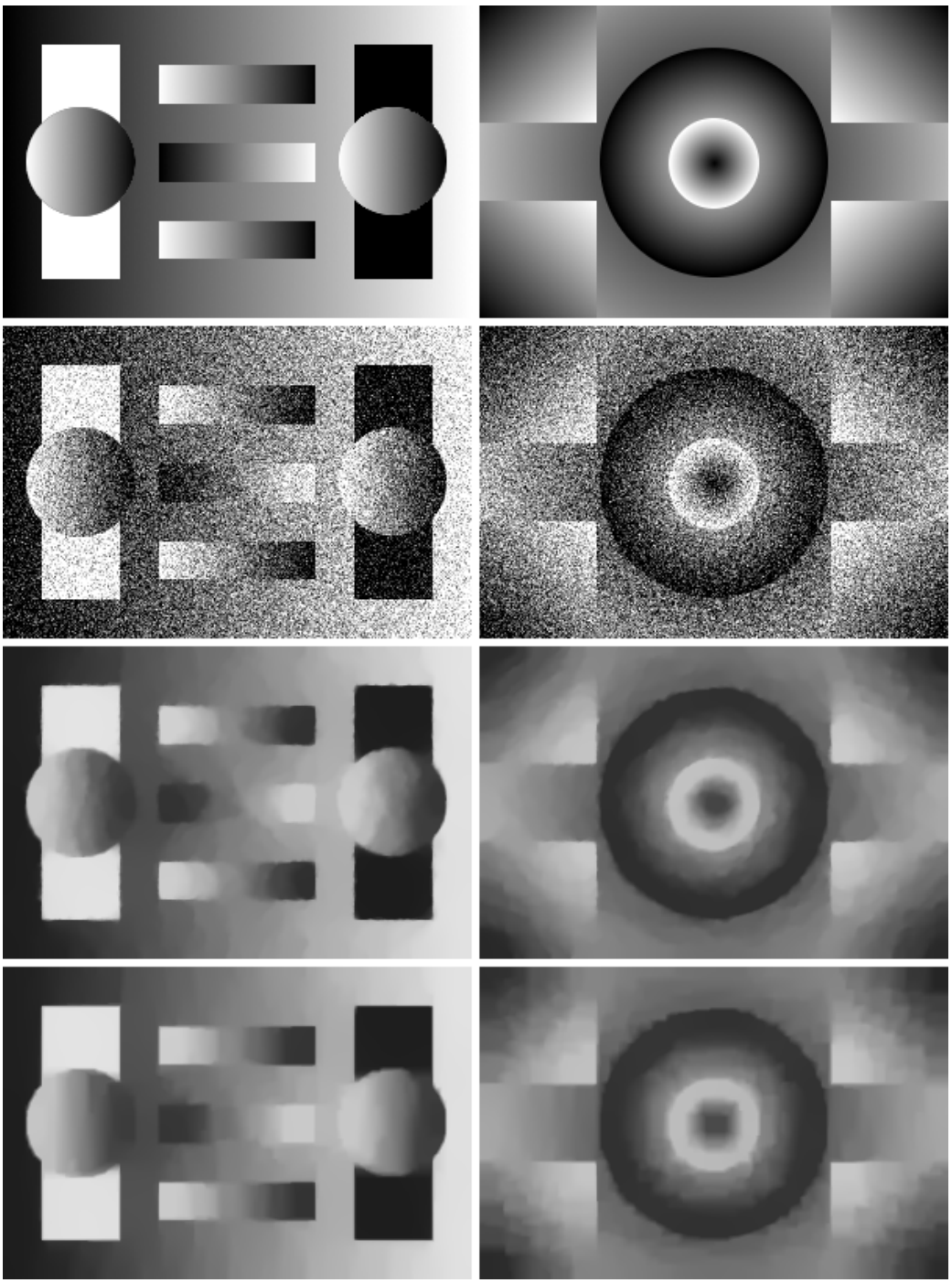
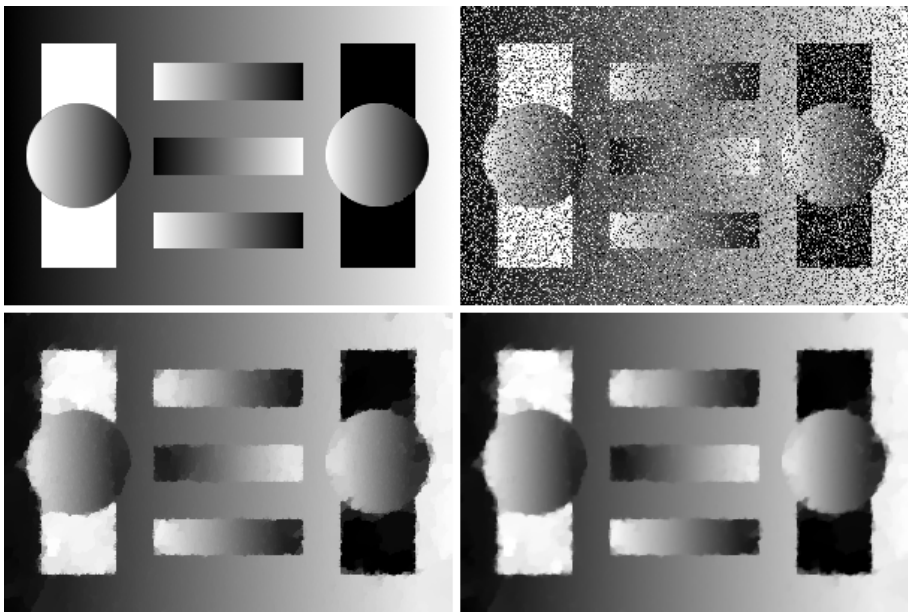


Fig. 11 Isotropic versus anisotropic second order denoising. The value of parameters was $\alpha = 0.3$, $\beta = 0.07$, number of iterations was 200, for all the images. First row: Original images. Second row: Images corrupted with Gaussian noise of variance 0.05. Third row: Isotropic second order denoising (SSIM=0.8153, 0.8612 from left to right). Fourth row: Anisotropic second order denoising (SSIM=0.8197, 0.8445 from left to right). It is clear that in parts of the image where the gradient is parallel or vertical to the x-axis, the anisotropic denoising performs better. This is not the case for general directions of the gradient where the isotropic method produces better results

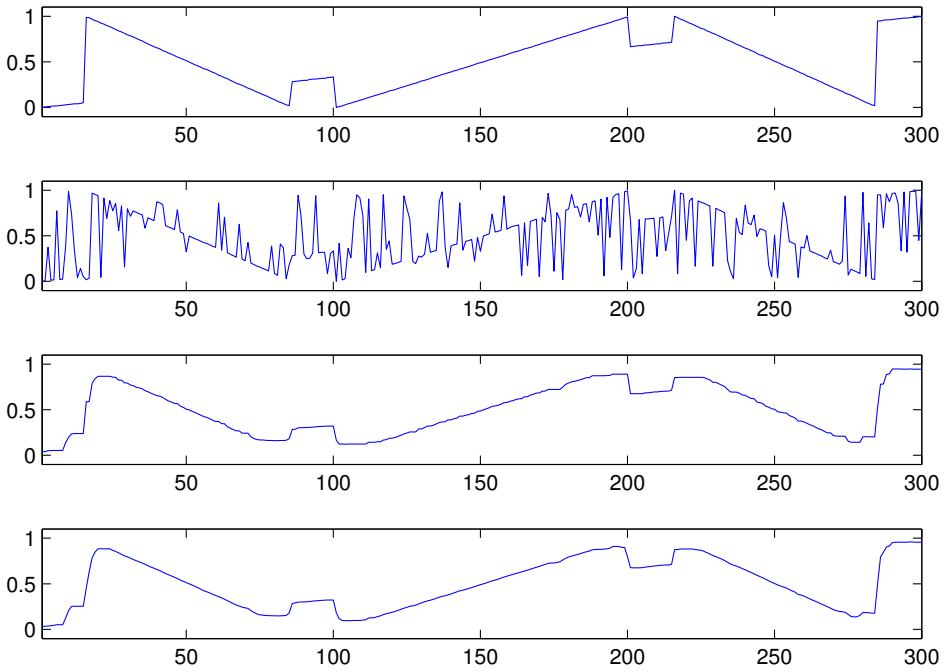
separately. As it is seen in the zoomed parts in the last row's subfigures, the second order method outperforms TV, in terms of the connectivity principle. Note also that in the second order inpainting, we used a small value of $\alpha = 10^{-6}$. The result looks similar if we choose $\alpha = 0$.

Computational times

All the numerical computations were done on a MAC OS X 3.06 GHz, 4 GB RAM using MATLAB. In Table 1 we display the computational times in order to achieve a relative residual $\|u_k - u_{k-1}\|/\|u_k\| < 10^{-4}$ for the examples of Figure 5. We compare the computational times of the two splitting techniques that were discussed in Section 5.3. We notice that relative resid-



(a) (u.l.) clean image, (u.r.) image corrupted with impulse noise of density 0.4 (SSIM=0.0880), (b.l) restored image with $\alpha = 1.4$, $\beta = 0$ (SSIM=0.8717), (b.r.) restored image with $\alpha = 0.9$, $\beta = 0.24$ (SSIM=0.8817)



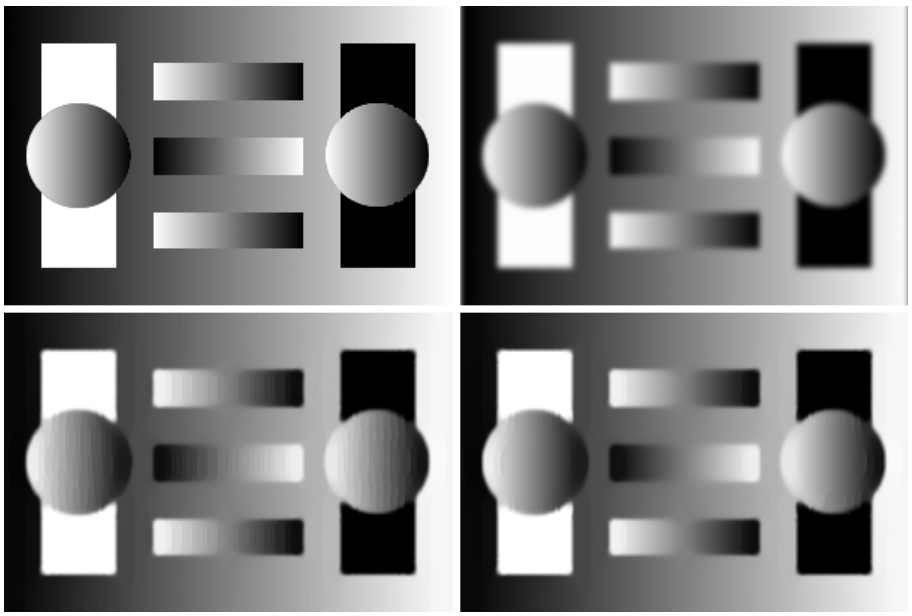
(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (u.l), (u.r), (b.l) and (b.r.) images respectively

Fig. 12 Impulse noise denoising with L^1 norm in the fidelity term. Image corrupted with impulse noise of density 0.4. Number of iterations=500

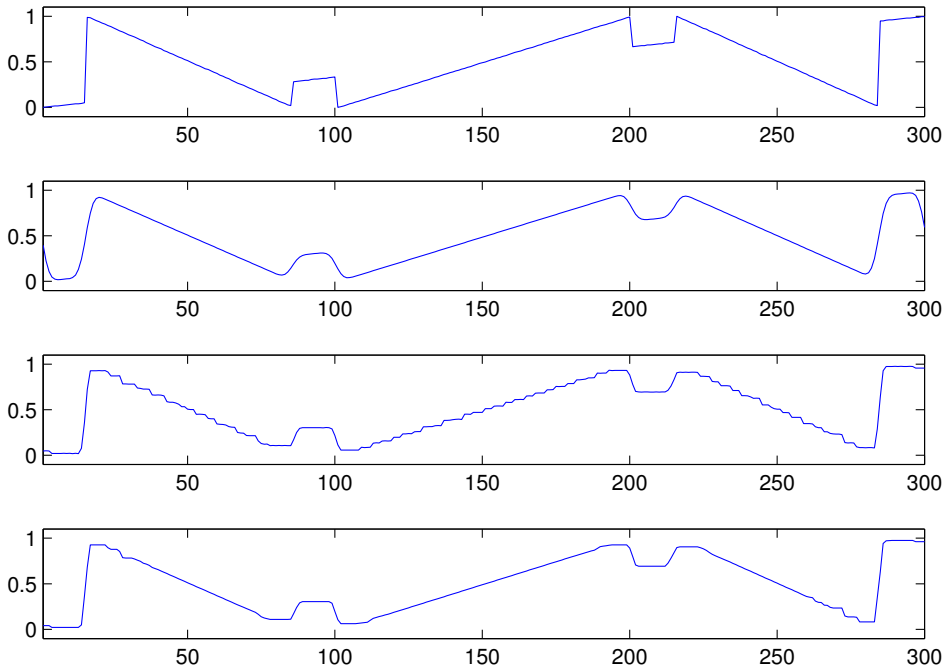
ual of order 10^{-4} is achieved in at most 60 iterations. The fastest time for the second order method is 3.73 seconds versus 2.66 seconds for TV-minimisation and it is achieved using 4 iterations of a preconditioned Conjugate Gradient method (*pcg* function in MATLAB) in order to solve the linear system in each iteration. The alternative splitting did not seem to provide faster convergence. However we believe that we can achieve faster computational times by optimizing further our code , e.g., implementation in C++, alternative linear system solver. This is a work in progress.

9 Conclusion

We formulate a second order variational problem in the space of functions of bounded Hessian in the context of convex functions of measures. We prove exis-



(a) (u.l.) clean image, (u.r.) image convoluted with Gaussian kernel of variance $\sigma = 2$ (SSIM=0.8659), (b.l) restored image with $\alpha = 0.06$, $\beta = 0$ (SSIM=0.9066), (b.r.) restored image with $\alpha = 0.06$, $\beta = 0.004$ (SSIM=0.9173)

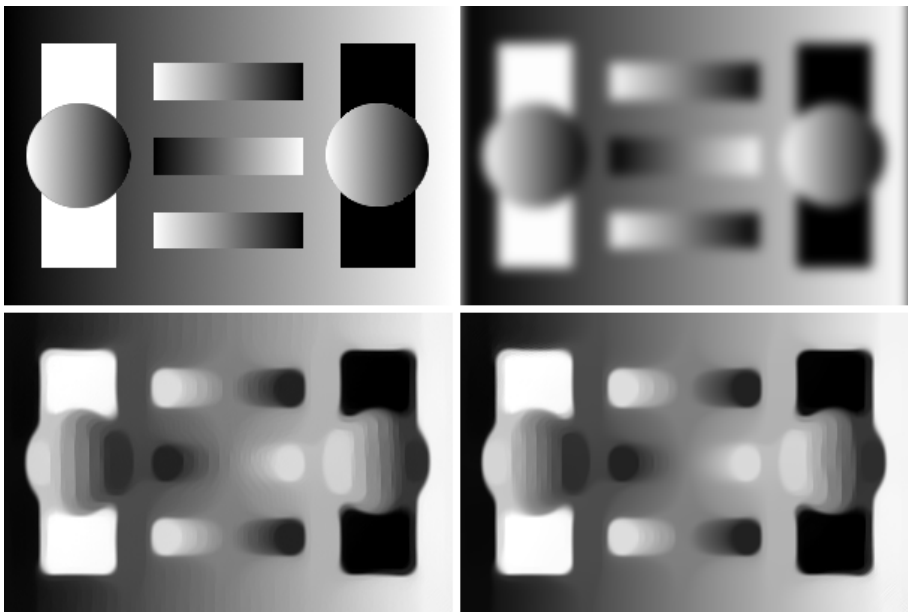


(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (u.l), (u.r), (b.l) and (b.r.) images respectively

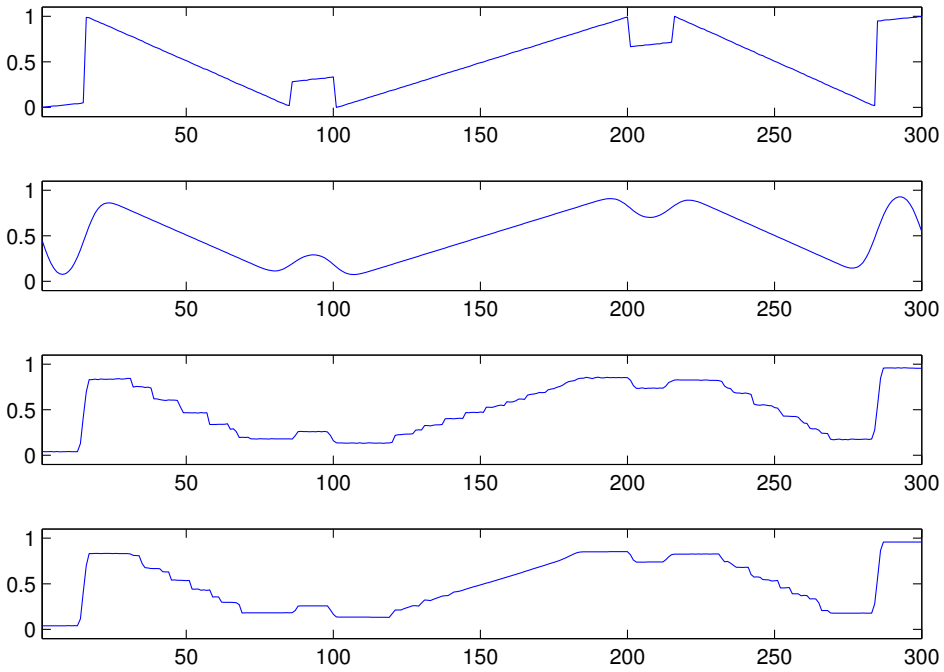
Fig. 13 Deblurring. Variance of the convolution kernel is $\sigma = 2$. Number of iterations = 100. A relatively small value of β can reduce significantly the staircasing effect

tence and uniqueness of minimisers using a relaxation technique. We propose the use of the split Bregman method for the numerical solution of the analogue discretised problem. The application of split Bregman method in our model is quite robust and is converging after a few iterations. We perform numerical experiments in denoising, reconstructing images that have been corrupted with Gaussian and impulse noise, in deblurring deconvoluting images that have been convoluted with Gaussian kernels, as well as in image inpainting.

In every case, the introduction of the second order term leads to a significant reduction of the staircasing effect resulting more in piecewise smooth images rather than piecewise constant as in images reconstructed using the ROF model. The superiority of



(a) (u.l.) clean image, (u.r.) image convoluted with Gaussian kernel of variance $\sigma = 4$ (SSIM=0.7392), (b.l) restored image with $\alpha = 0.18$, $\beta = 0$ (SSIM=0.8105), (b.r.) restored image with $\alpha = 0.18$, $\beta = 0.002$ (SSIM=0.8129)

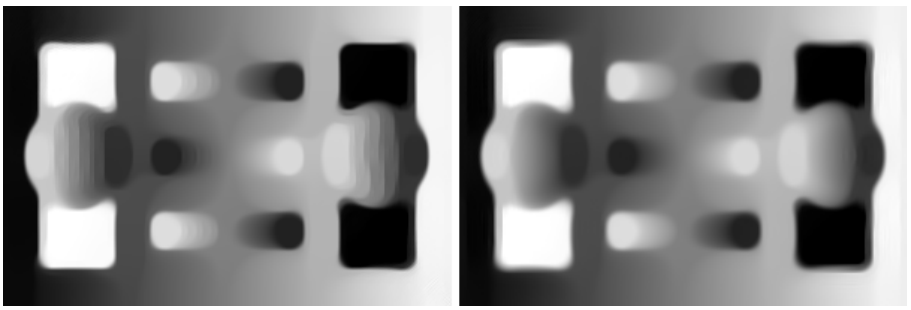


(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (u.l), (u.r), (b.l) and (b.r.) images respectively

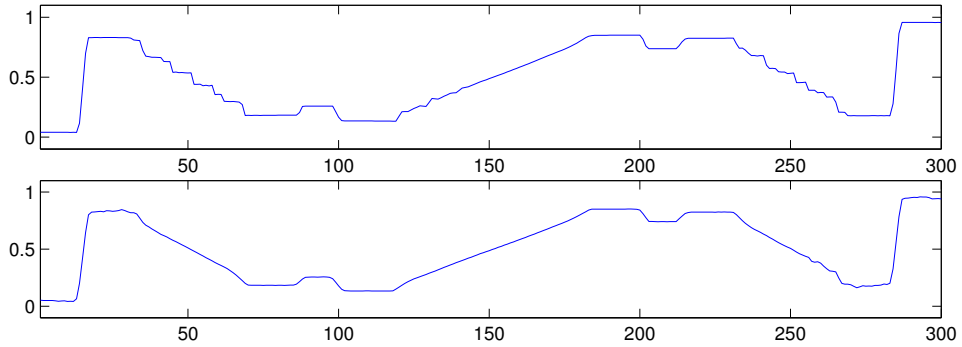
Fig. 14 Deblurring. Variance of the convolution kernel is $\sigma = 4$. Number of iterations = 100. A relatively small value of β reduces the staircasing effect

using a combination of the first and the second order rather than the first order only, is confirmed quantitatively by the SSIM index. In the case of inpainting the higher order able is able to connect edges along large gaps something that TV-inpainting is incapable of.

As fas as future work is concerned, there is some room for improvement concerning the optimality of the programming code, as our implementation is relatively slow, especially in deblurring. Moreover, the relation between the continuum and the discretised model could be investigated through Γ-convergence arguments, see [19] and [7]. Moreover, the characterisation of subgradients of this approach and the analysis of solutions of the corresponding PDE flows for different choices of functions f and g promises to give



(a) (l.) restored image with best SSIM $\alpha = 0.18$, $\beta = 0.002$ (SSIM=0.8129), (r.) restored image with higher β, $\alpha = 0.18$, $\beta = 0.004$ (SSIM=0.7894)



(b) Middle row slices of the above pictures. From top to bottom, the slices correspond to (l.) and (r.) images respectively

Fig. 15 Further reduction of the staircasing effect by increasing the value of β. Variance of the Gaussian convolution kernel is $\sigma = 4$

Method	Splitting	Lin. System Solver	Seconds	No. of Iterations	Rel. Error
TV	standard	LU	6.49	54	0.28 %
TV	standard	PCG	**2.66**	51	0.26 %
TV-BH	standard	LU	7.53	47	0.20 %
TV-BH	standard	PCG	**3.73**	37	0.33 %
TV-BH	alternative	LU	20.07	61	0.31%
TV-BH	alternative	PCG	10.77	58	0.33 %

Table 1 Computational times for the two examples of Figure 5 to reach a relative residual $\|u_k - u_{k-1}\|/\|u_k\| < 10^{-4}$. The linear system solver is either LU-Factorisation or a Preconditioned Conjugate Gradient method (MATLAB's built in function) with stopping criteria either maximum 4 iterations or residual less than 10^{-4}. The last column of the table shows the relative error between the results and the "correct" solution, where the correct solutions are chosen the ones that are shown in Figure 5 (300 iterations)

more insight into the qualitative properties of this regularisation procedure. Higher order models in image reconstruction form currently an active field of research. During the last years many methods have been introduced, the most important of which are mentioned in the Section 1.2. It will be interesting to compare these methods with ours in particular with the new notion of total generalised variation TGV which has produced some very good results, see [8].

Acknowledgements The authors acknowledge the financial support provided by the Cambridge Centre for Analysis (CCA) and the Royal Society International Exchanges Award IE110314 for the project "High-order Compressed Sensing for Medical Imaging". Further, this publication is based on work supported by Award No. KUK-I1-007-43 , made by King Abdullah University of Science and Technology (KAUST).

A Some results about convex functions of measures

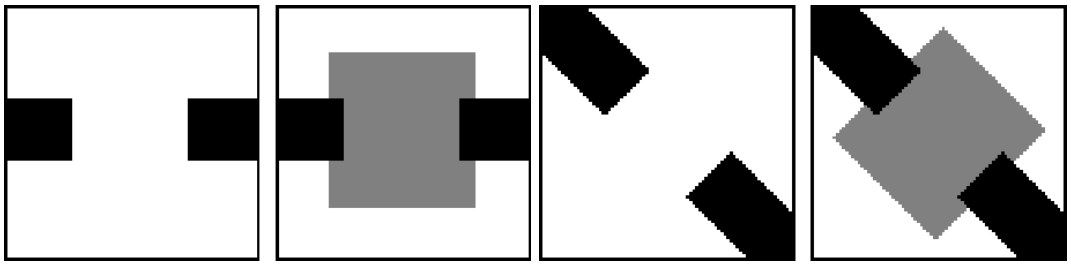
Proposition A.1. *Suppose that $g : \mathbb{R}^m \to \mathbb{R}$ is continuous function, positively homogeneous of degree 1 and let $\mu \in [\mathcal{M}(\Omega)]^m$. Then for every positive measure Radon measure ν such that μ is absolulely continuous with respect to ν, we have*

$$g(\mu) := g\left(\frac{\mu}{|\mu|}\right)|\mu| = g\left(\frac{\mu}{\nu}\right)\nu.$$

Moreover, if g is a convex function, then $g : [\mathcal{M}(\Omega)]^m \to \mathcal{M}(\Omega)$ is a convex function as well.

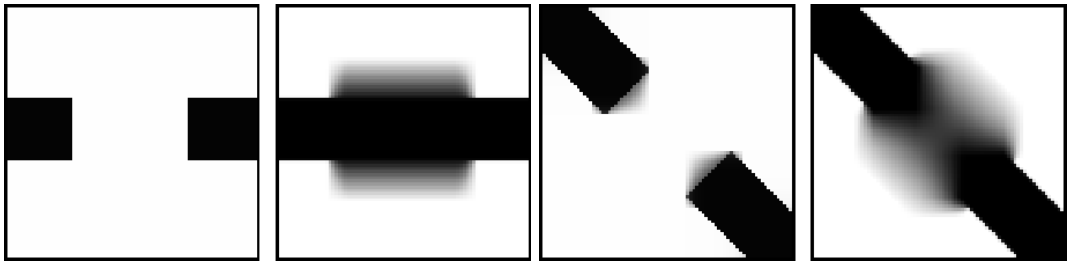
Proof. Since $\mu \ll \nu$, we have that $|\mu| \ll \nu$. Using the fact that g is positively homogenous we get

$$g(\mu) = g\left(\frac{\mu}{|\mu|}\right)|\mu| = g\left(\frac{\mu}{|\mu|}\right)\frac{|\mu|}{\nu}\nu = g\left(\frac{\mu}{|\mu|}\frac{|\mu|}{\nu}\right)v = g\left(\frac{\mu}{\nu}\right)\nu.$$



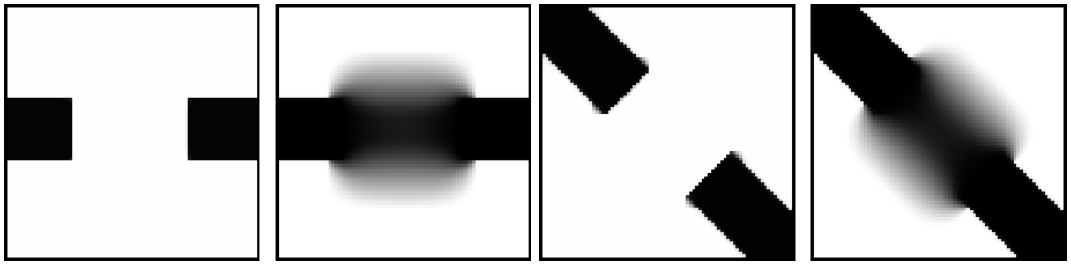
(a) (l.) Original image, (r.) Inpainting domain (grey)

(b) (l.) Original image, 45° tilted version of (a), (r.) Inpainting domain (grey)



(c) (l.) Anisotropic TV-inpainting, $\alpha = 0.12$, $\beta = 0$, 4000 iterations, (r.) Anisotropic second order inpainting, $\alpha = 10^{-4}$, $\beta = 0.01$, 1500 iterations

(d) (l.) Anisotropic TV-inpainting, $\alpha = 0.12$, $\beta = 0$, 4000 iterations, (r.) Anisotropic second order inpainting, $\alpha = 10^{-4}$, $\beta = 0.01$, 1500 iterations



(e) (l.) Isotropic TV-inpainting, $\alpha = 0.12$, $\beta = 0$, 4000 iterations, (r.) Isotropic second order inpainting, $\alpha = 10^{-4}$, $\beta = 0.01$, 1500 iterations

(f) (l.) Isotropic TV-inpainting, $\alpha = 0.12$, $\beta = 0$, 4000 iterations, (r.) Isotropic second order inpainting, $\alpha = 10^{-4}$, $\beta = 0.01$, 1500 iterations

Fig. 16 Inpainting geometrical images

Assuming that g is convex and using the first part of the proposition we get for $0 \leq \lambda \leq 1$, μ, $\nu \in [\mathcal{M}(\Omega)]^m$:

$$g(\lambda\mu + (1-\lambda)\nu) = g\left(\frac{\lambda\mu + (1-\lambda)\nu}{|\lambda\mu + (1-\lambda)\nu|}\right)|\lambda\mu + (1-\lambda)\nu|$$

$$= g\left(\frac{\lambda\mu + (1-\lambda)\nu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|)$$

$$= g\left(\lambda\frac{\mu}{|\mu| + |\nu|} + (1-\lambda)\frac{\nu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|)$$

$$\leq \lambda g\left(\frac{\mu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|) + (1-\lambda)g\left(\frac{\nu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|)$$

$$= \lambda g\left(\frac{\mu}{|\mu|}\right)|\mu| + (1-\lambda)g\left(\frac{\nu}{|\nu|}\right)|\nu|$$

$$= \lambda g(\mu) + (1-\lambda)g(\nu).$$

□

The following theorem which is a special case of a theorem that was proved in [11] and can be also found in [2] establishes the lower semicontinuity of convex functionals of measures with respect to the weak* convergence.

Theorem A.2 (Buttazzo-Freddi, 1991)**.** *Let Ω be an open subset of $\mathbb{R}^n$, ν, $(\nu_k)_{k\in\mathbb{N}}$ be $\mathbb{R}^m$-valued finite Radon measures and μ, $(\mu_k)_{k\in\mathbb{N}}$ be positive Radon measures in Ω. Let $g : \mathbb{R}^m \to \mathbb{R}$ be a convex function and suppose that $\nu_k \to \nu$ and $\mu_k \to \mu$ weakly* in Ω. Consider the Lebesgue decompositions $\nu = (\nu/\mu)\mu + \nu^s$, $\nu_k = (\nu_k/\mu_k)\mu_k + \nu_k^s$, $k \in \mathbb{N}$. Then*

$$\int_\Omega g\left(\frac{\nu}{\mu}(x)\right)d\mu(x) + \int_\Omega g_\infty\left(\frac{\nu^s}{|\nu^s|}(x)\right)d|\nu^s|(x) \leq$$

$$\liminf_{k\to\infty}\int_\Omega g\left(\frac{\nu_k}{\mu_k}(x)\right)d\mu_k(x) + \int_\Omega g_\infty\left(\frac{\nu_k^s}{|\nu_k^s|}(x)\right)d|\nu_k^s|(x).$$

In particular, if $\mu = \mu_k = \mathcal{L}^n$ for all $k \in \mathbb{N}$ then according to the definition (2.1) *the above inequality can be written as follows:*

$$g(\nu)(\Omega) \leq \liminf_{k\to\infty} g(\nu_k)(\Omega).$$

The following theorem is a special case of Theorem 2.3 in [22].

Theorem A.3 (Demengel-Temam, 1984)**.** *Suppose that $\Omega \subseteq \mathbb{R}^n$ is open, with Lipschitz boundary and let g be a convex function from $\mathbb{R}^{n\times n}$ to $\mathbb{R}$ with at most linear growth at infinity. Then for every $u \in BH(\Omega)$ there exists a sequence $(u_k)_{k\in\mathbb{N}} \subseteq C^\infty(\Omega) \cap W^{2,1}(\Omega)$ such that*

$$\|u_k - u\|_{L^1(\Omega)} \to 0, \quad |D^2u_k|(\Omega) \to |D^2u|(\Omega),$$

$$g(D^2u_k)(\Omega) \to g(D^2u)(\Omega), \quad \text{as} \quad k \to \infty.$$

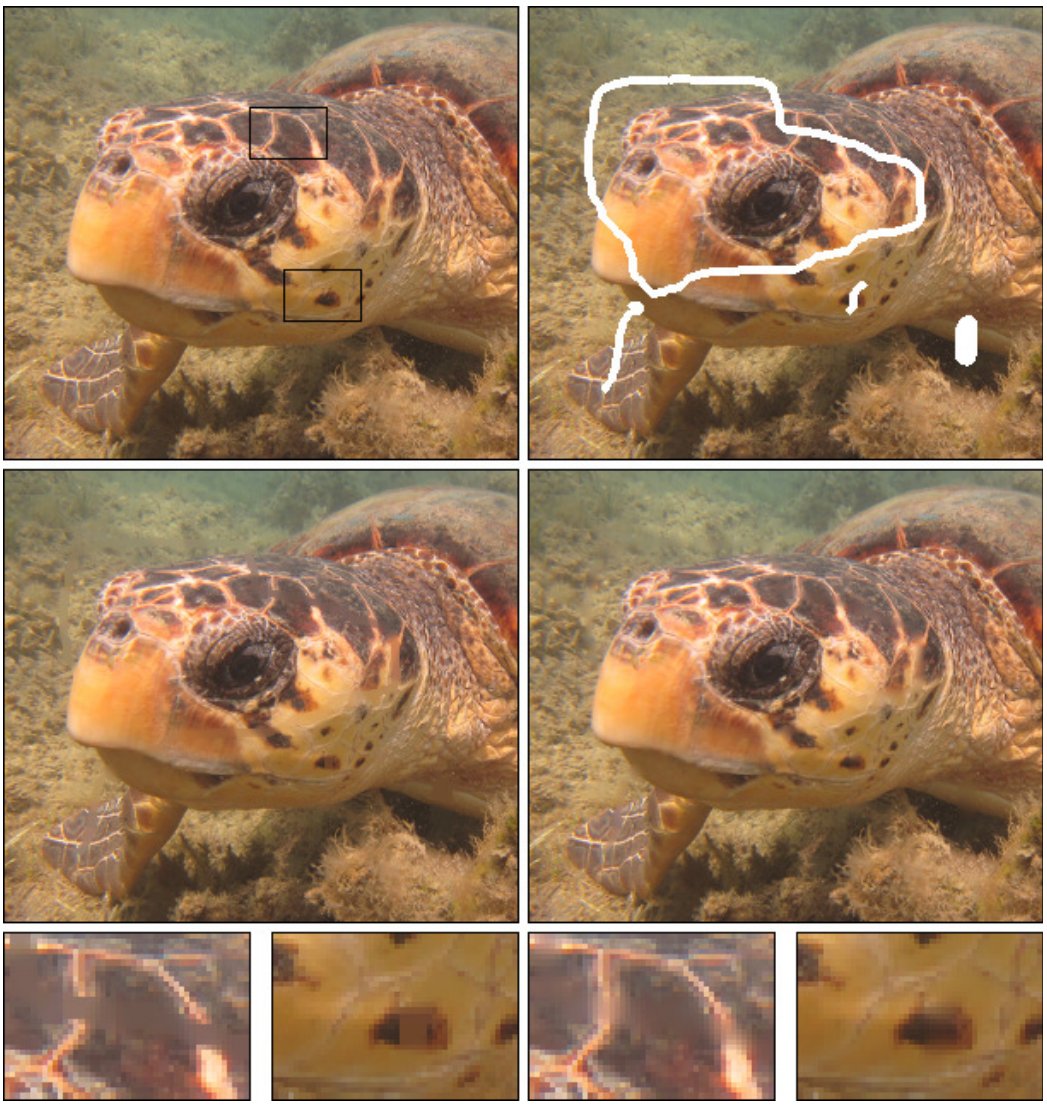


Fig. 17 Inpainting a real world image: *Sotiris the turtle.* (u.l) original image, (u.r.) destroyed image and initialisation of the algorithm, (m.l) TV-inpainting, $\alpha = 0.001$, $\beta = 0$, 7000 iterations, (m.r.) second order inpainting, $\alpha = 10^{-6}$, $\beta = 0.0005$, 5000 iterations. In the last rows we have enlarged some details of the two reconstructed images

B Weak continuity and differentiability notions in *BV*

The purpose of this appendix is to remind the reader of the notions of approximate limits, continuity, jump points and differentiability of BV functions. We remind also the decomposition of the distributional derivative into an absolutely continuous part, a jump part and a Cantor part. We assume that $\Omega \subseteq \mathbb{R}^n$ is open with Lipschitz boundary.

Definition B.1 (Approximate limit-approximate continuity)**.** *Let $u \in [L^1(\Omega)]^m$. We say that u has an approximate limit at $x \in \Omega$ if there exists a $z \in \mathbb{R}^m$ such that*

$$\lim_{\rho \to 0} \frac{1}{|B_\rho(x)|} \int_{B_\rho(x)} |u(y) - z| dy = 0.$$

We denote with S_u the set of points where this property does not hold. The vector z, which is unique if it exists, is denoted by $\tilde{u}(x)$. If $\tilde{u}(x) = u(x)$ we say that u is approximately continuous in x .

Definition B.2 (Approximate jump point)**.** *We say that x is an approximate jump point if there exist a, $b \in \mathbb{R}^m$ and $\nu \in \mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \ |x| = 1\}$ such that $a \neq b$ and*

$$\lim_{\rho \to 0} \frac{1}{|B^+_\rho(x,\nu)|} \int_{B^+_\rho(x,\nu)} |u(y) - a| \ dy = 0,$$

$$\lim_{\rho \to 0} \frac{1}{|B^-_\rho(x,\nu)|} \int_{B^-_\rho(x,\nu)} |u(y) - b| \ dy = 0,$$

with

$$B^+_\rho(x) = \{y \in B_\rho(x) : \ (y - x, \nu) > 0\},$$

$$B^-_\rho(x) = \{y \in B_\rho(x) : \ (y - x, \nu) < 0\},$$

where $(\cdot, \cdot)$ denotes the Euclidean inner product. The set of all approximate jump points is denoted by J_u. The triplet (a, b, ν) will be denoted by $(u^+(x), u^-(x), \nu_u(x))$.

Definition B.3 (Approximate differentiability)**.** *Let $u \in [L^1(\Omega)]^m$. We say that u is approximately differentiable at a point $x \in \Omega \setminus S_u$ if there is a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{\rho \to 0} \frac{1}{|B_\rho(x)|} \int_{B_\rho(x)} \frac{|u(y) - \tilde{u}(x) - L(y - x)|}{\rho} \ dy = 0.$$

The linear functional L is represented by an $m \times n$ matrix which is called the approximate differential of f at x and will be denoted by $\nabla u(x)$.

Note that the approximate differential is a linear function of u. It is a fact that if $u \in [BV(\Omega)]^m$ then $\mathcal{H}^{n-1}(S_u \setminus J_u) = 0$. Moreover, Calderón-Zygmund theorem says that

u is approximately differentiable at $\mathcal{L}^n$-almost every point of Ω and the approximate differential ∇u is the density of the absolutely continuous part of Du with respect to $\mathcal{L}^n$ i.e. $Du = \nabla u \mathcal{L}^n + D^s u$. Consequently we have the following decomposition for the derivative of u:

$$Du = D^a u + D^s u = \nabla u \mathcal{L}^n + D^j u + D^c u,$$

where $D^j u := D^s u \restriction J_u$, the restriction of $D^s u$ on J_u, and $D^c u := D^s u \restriction (\Omega \setminus S_u)$ are the *jump* and the *Cantor* part of the derivative respectively. Moreover it can be shown that the jump part is equal to

$$D^j(B) = \int_{B \cap J_u} (u^+(x) - u^-(x)) \otimes \nu_u(x)\, d\mathcal{H}^{n-1}(x),\ \forall B \in \mathcal{B}(\Omega).$$

We recall that if $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ then $a \otimes b$ denotes the $m \times n$ matrix with (ij)-th entry $a_i b_j$.This implies that

$$|Du|(\Omega) = \int_\Omega |\nabla u(x)| dx + \int_{J_u} |u^+(x) - u^-(x)|\, d\mathcal{H}^{n-1}(x) + |D^c u|(\Omega \setminus S_u). \tag{B.1}$$

C Formulation of the discrete second divergence operators

In this appendix we define the first and second discrete divergence operators div and divH, that have the properties:

$$\mathrm{div} : \left(\mathbb{R}^{n\times m}\right)^2 \to \mathbb{R}^{n\times m} \quad \text{with} \quad -\mathrm{div}(v) \cdot u = v \cdot \nabla u,$$

$$\forall u \in \mathbb{R}^{n\times m},\ v \in \left(\mathbb{R}^{n\times m}\right)^2,$$

$$\mathrm{divH} : \left(\mathbb{R}^{n\times m}\right)^4 \to \mathbb{R}^{n\times m} \quad \text{with} \quad \mathrm{divH}(w) \cdot u = w \cdot Hu,$$

$$\forall u \in \mathbb{R}^{n\times m},\ w \in \left(\mathbb{R}^{n\times m}\right)^4.$$

The formulation we mention here is due to [5]. For $v = (v_1, v_2) \in (\mathbb{R}^{n\times m})^2$ we have that $\mathrm{div}(v) = \mathrm{div}_1(v_1) + \mathrm{div}_2(v_2)$ where $\mathrm{div}_1(v_1)$, $\mathrm{div}_2(v_2)$ belong to $\mathbb{R}^{n\times m}$ and

$$\mathrm{div}_1(v_1)(i,j) = \begin{cases} v_1(i,j) - v_1(i,j-1) & \text{if } 1 < j < m, \\ v_1(i,j) & \text{if } j = 1, \\ -v_1(i,j-1) & \text{if } j = m, \end{cases}$$

$$\mathrm{div}_2(v_2)(i,j) = \begin{cases} v_2(i,j) - v_2(i-1,j) & \text{if } 1 < i < n, \\ v_2(i,j) & \text{if } i = 1, \\ -v_2(i-1,j) & \text{if } i = n. \end{cases}$$

For the discrete second divergence divH we have that for $w = (w_{11}, w_{12}, w_{21}, w_{22}) \in (\mathbb{R}^{n\times m})^4$ we have that $\mathrm{divH}(w) = \mathrm{divH}_{11}(w_{11}) + \mathrm{divH}_{12}(w_{12}) + \mathrm{divH}_{21}(w_{21}) + \mathrm{divH}_{22}(w_{22})$ where $\mathrm{divH}_{11}(w_{11})$, $\mathrm{divH}_{12}(w_{12})$, $\mathrm{divH}_{21}(w_{21})$, $\mathrm{divH}_{22}(w_{22})$ belong to $\mathbb{R}^{n\times m}$ and

$$\mathrm{divH}_{11}(w_{11})(i,j) = \begin{cases} w_{11}(i,j-1) - 2w_{11}(i,j) & \text{if } 1 < j < m, \\ +w_{11}(i,j+1) & \\ w_{11}(i,j+1) - w_{11}(i,j) & \text{if } j = 1, \\ w_{11}(i,j-1) - w_{11}(i,j) & \text{if } i = m, \end{cases}$$

$$\mathrm{divH}_{22}(w_{22})(i,j) = \begin{cases} w_{22}(i-1,j) - 2w_{22}(i,j) & \text{if } 1 < i < n, \\ +w_{22}(i+1,j) & \\ w_{22}(i+1,j) - w_{22}(i,j) & \text{if } i = 1, \\ w_{22}(i-1,j) - w_{22}(i,j) & \text{if } i = n, \end{cases}$$

$$\mathrm{divH}_{12}(w_{12})(i,j) = \begin{cases} w_{12}(i-1,j) - w_{12}(i,j) & \text{if } 1 < i, \\ -w_{12}(i-1,j+1) & j < m \\ +w_{12}(i,j+1) & \\ w_{12}(i,j+1) - w_{12}(i,j) & \text{if } i = 1, \\ & 1 < j < m, \\ w_{12}(i-1,j) & \text{if } i = n, \\ -w_{12}(i-1,j+1) & 1 < j < m, \\ w_{12}(i,j+1) & \text{if } 1 < i < n, \\ -w_{12}(i-1,j+1) & j = 1, \\ w_{12}(i-1,j) - w_{12}(i,j) & \text{if } 1 < i < n, \\ & j = m, \\ w_{12}(i,j+1) & \text{if } i = 1, \\ & j = 1, \\ -w_{12}(i,j) & \text{if } i = 1, \\ & j = m, \\ -w_{12}(i-1,j+1) & \text{if } i = n, \\ & j = 1 \\ w_{12}(i-1,j) & \text{if } i = n, \\ & j = m, \end{cases}$$

$$\mathrm{divH}_{21}(w_{21})(i,j) = \begin{cases} w_{21}(i,j-1) - w_{21}(i,j) & \text{if } 1 < i, \\ -w_{21}(i+1,j-1) & j < m, \\ +w_{21}(i+1,j) & \\ w_{21}(i+1,j) & \text{if } i = 1, \\ -w_{21}(i+1,j-1) & 1 < j < m, \\ w_{21}(i,j-1) - w_{21}(i,j) & \text{if } i = n, \\ & 1 < j < m, \\ w_{21}(i+1,j) - w_{21}(i,j) & \text{if } 1 < i < n, \\ & j = 1 \\ w_{21}(i,j-1) & \text{if } 1 < i < n, \\ -w_{21}(i+1,j-1) & j = m, \\ w_{21}(i+1,j) & \text{if } i = 1, \\ & j = 1, \\ -w_{21}(i+1,j-1) & \text{if } i = 1, \\ & j = m, \\ -w_{21}(i,j) & \text{if } i = n, \\ & j = 1 \\ w_{21}(i,j-1) & \text{if } i = n, \\ & j = m. \end{cases}$$

D Some useful theorems

Lemma D.1 (Kronecker's lemma). *Suppose that $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are two sequences of real numbers such that $\sum_{n=1}^\infty a_n < \infty$ and $0 < b_1 \le b_2 \le \ldots$ with $b_n \to \infty$. Then*

$$\frac{1}{b_n} \sum_{k=1}^n b_k a_k \to 0, \quad \text{as } n \to \infty.$$

In particular, if $(c_n)_{n\in\mathbb{N}}$ is a decreasing positive real sequence such that $\sum_{n=1}^\infty c_n^2 < \infty$, then

$$c_n \sum_{k=1}^n c_k \to 0, \quad \text{as } n \to \infty.$$

References

1. R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. In *Inverse Problems*, pages 1217–1229, 1994.

2. L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford University Press, USA, 2000.
3. M. Benning. *Singular Regularization of Inverse Problems*. 2011.
4. M. Benning, C. Brune, M. Burger, and J. Müller. Higher-order TV methods -Enhancement via Bregman iteration. *UCLA CAM Report 12-04*, 2012.
5. M. Bergounioux and L. Piffet. A second-order model for image denoising. *Set Valued Anal. Variational Anal*, 18(3-4):277–306, 2010.
6. A. L. Bertozzi and J. B. Greer. Low-curvature image simplifiers: Global regularity of smooth solutions and laplacian limiting schemes. *Communications on Pure and Applied Mathematics*, 57(6):764–790, 2004.
7. A. Braides. *Γ-convergence for beginners*. Oxford lecture series in mathematics and its applications, 2002.
8. K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3:1–42, 2009.
9. K. Bredies and T. Valkonen. Inverse problems with second-order total generalized variation constraints. In *Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications, Singapore*, 2011.
10. L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.
11. G. Buttazzo and L. Freddi. Functionals defined on measures and applications to non equi-uniformly elliptic problems. *Annali di Matematica Pura ed Applicata*, 159(1):133–149, 1991.
12. A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.
13. A. Chambolle and P. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.
14. T. Chan and S. Esedoglu. Aspects of total variation regularized L^1 function approximation. *SIAM Journal on Applied Mathematics*, pages 1817–1837, 2005.
15. T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM J. Sci. Comput.*, 22(2):503–516, 2001.
16. T. F. Chan, E. S., and F. Park. Image decomposition combining staircase reduction and texture extraction. *J. Visual Communication and Image Representation*, 18(6):464–486, 2007.
17. P. Charbonnier, L. Blanc-Fraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *ICIP (2)'94*, pages 168–172, 1994.
18. P. Combettes, V. Wajs, et al. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4):1168–1200, 2006.
19. G. Dal Maso. *Introduction to Γ-convergence*. Birkhauser, 1993.
20. G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini. A higher order model for image restoration: the one dimensional case. *SIAM J. Math. Anal*, 40(6):2351–2391, 2009.
21. F. Demengel. Fonctions à hessian borné. *Ann. Inst. Fourier*, 34:155–190, 1985.
22. F. Demengel and R. Temam. Convex functions of a measure and applications. *Indiana Univ. Math. J.*, 33:673–709, 1984.
23. S. Didas, J. Weickert, and B. Burgeth. Properties of higher order nonlinear diffusion filtering. *J. Math. Imaging Vis.*, 35:208–226, 2009.
24. Y. Dong, M. Hintermüller, F. Knoll, and R. Stollberger. Total variation denoising with spatially dependent regularization. In *ISMRM 18th Annual Scientific Meeting and Exhibition Proceedings*, page 5088, 2010.
25. V. Duval, J. Aujol, and Y. Gousseau. The TV$L1$ model: a geometric point of view. *SIAM journal on multiscale modeling and simulation*, 8(1):154–189, 2009.
26. E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for tv minimization. *UCLA CAM Report 09-67*, 2009.
27. L. Evans. *Partial Differential Equations, volume 19 of Graduate Studies in Mathematics, Second Edition*. American Mathematical Society,, 2010.
28. X. Feng and A. Prohl. Analysis of total variation flow and its finite element approximations. *ESAIM-Mathematical Modelling and Numerical Analysis*, 37(3):533, 2003.
29. K. Frick, P. Marnitz, and A. Munk. Statistical multiresolution estimation in imaging: Fundamental concepts and algorithmic framework. *Arxiv preprint arXiv:1101.4373*, 2011.
30. C. Frohn-Schauf, S. Henn, and K. Witsch. Nonlinear multigrid methods for total variation image denoising. *Computing and Visualization in Science*, 7(3):199–206, 2004.
31. T. Goldstein and S. Osher. The split bregman method for $L1$-regularized problems. *SIAM Journal on Imaging Sciences*, 2:323, 2009.
32. P. J. Green. Bayesian reconstructions from emission tomography data using a modified em algorithm. *IEEE Trans. Med. Imag*, pages 84–93, 1990.
33. W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded hessian for convexification and denoising. *Computing*, 76:109–133, 2006.
34. R. Lai, X. Tai, and T. Chan. A ridge and corner preserving model for surface restoration. *UCLA CAM Report 11-55*, 2011.
35. M. Lysaker, A. Lundervold, and X. C. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, 2003.
36. M. Lysaker and X. C. Tai. Iterative image restoration combining total variation minimization and a second-order functional. *International Journal of Computer Vision*, 66(1):5–18, 2006.
37. M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers. *SIAM Journal on Numerical Analysis*, pages 965–994, 2003.
38. M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1):99–120, 2004.
39. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4:460–489, 2005.
40. C. Pöschl and O. Scherzer. Characterization of minimisers of convex regularization functionals. *Contemporary mathematics*, 451:219–248, 2008.
41. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.
42. O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *jour-CASC*, 60(1):1–27, 1998.
43. S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. *Scale Space and Variational Methods in Computer Vision*, 5567:464–476, 2009.
44. S. Setzer and G. Steidl. Variational methods with higher order derivatives in image processing. *Approximation XII*, pages 360–386, 2008.
45. S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete ℓ_1-type functionals. *Comm. Math. Sci*, 9:797–872, 2011.

46. A. Tikhonov. Solution of incorrectly formulated problems and the regularization method. In *Soviet Math. Dokl.*, volume 5, page 1035, 1963.
47. P. Vassilevski and J. G. Wade. A comparison of multilevel methods for total variation regularization. *Elec. Trans. Numer. Anal*, pages 6–255, 1997.
48. L. Vese. A study in the BV space of a denoising-deblurring variational problem. *Applied Mathematics and Optimization*, 44(2):131–161, 2001.
49. C. Vogel. A multigrid method for total variation-based image denoising. *Progress in Systems and Control Theory*, 20:323–323, 1995.
50. Y. Wang, W. Yin, and Y. Zhang. A fast algorithm for image deblurring with total variation regularization. *Image Rochester NY*, pages 1–19, 2007.
51. Z. Wang and A. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *Signal Processing Magazine, IEEE*, 26(1):98–117, 2009.
52. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.
53. E. T. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1923.
54. W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ_1-minimisation with applications to compressed sensing. *SIAM J. Imaging Science*, 1:142–168, 2008.